الأساليب الإحصائية

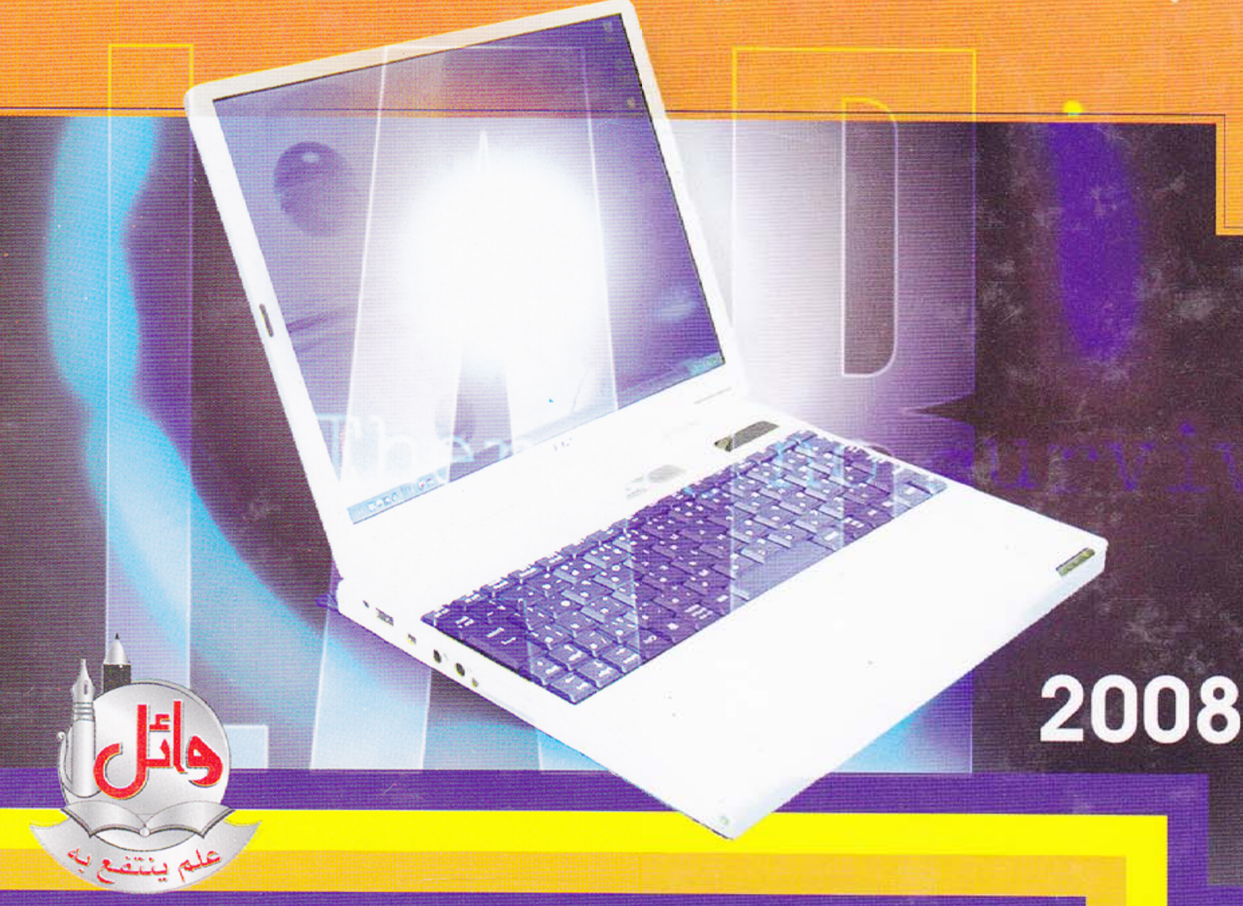
باستخدام حزمة

MATLAB

Programming Using MATLAB Package With
Statistical Applications

الأستاذ الدكتور
محمد عبد العال النعيمي
جامعة عمان العربية

الدكتور
مزهر شعبان العاني
جامعة عمان العربية



2008

الأساليب الاحصائية

باستخدام حزمة

MATLAB

Programming Using MATLAB Package with Statistical Applications

الدكتور

مزهر شعبان العاني

الأستاذ الدكتور

محمد عبد العال النعيمي

الاردن - عمان

دار وائل للنشر

الطبعة الأولى

2008

رقم الايداع لدى دائرة المكتبة الوطنية : (٢٠٥٠/٧/٢٠٠٧)
العاني ، مزهر شعبان
الأساليب الإحصائية باستخدام حزمة MATLAB / مزهر شعبان العاني، محمد عبد العال النعيمي . - عمان ، دار وائل ، ٢٠٠٧ .
(٢٥٧) ص
ر.إ. : (٢٠٥٠/٧/٢٠٠٧)
الواصفات: الإحصاء الوصفي / الحواسيب
* تم إعداد بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية
****

رقم التصنيف العشري / ديوي : ٩١٥.٥٣٠٢٨٥
ISBN 978-9957-11-702-3 (ردمك)

* الأساليب الاحصائية باستخدام حزمة MATLAB
* الدكتور مزهر شعبان العاني – الاستاذ الدكتور محمد عبد العال النعيمي
* الطبعـة الأولى ٢٠٠٨



دار وائــل للنشر والتوزيع

* الأردن – عمان – شارع الجمعية العلمية الملكية – مبنى الجامعة الاردنية الاستثماري رقم (٢) الطابق الثاني
هـاتف : ٠٠٩٦٢-٦-٥٣٣٨٤١٠ – فاكس : ٠٠٩٦٢-٦-٥٣٣١٦٦١ - ص. ب (١٦١٥ – الجبيهة)
* الأردن – عمـان – وسـط البـلد – مجمـع الفحيص التجــاري- هــاتف: ٠٠٩٦٢-٦-٤٦٢٧٦٢٧
www.darwael.com
E-Mail: Wael@Darwael.Com

الإهـــداء

إلى روح والدي وروح والدتي رحمهم الله عرفانا لكفاحهم

إلى أعزائي زوجتي وبناتي أيمانا لتضحيتهم

د. مزهر العاني

إلى روح والدي الطاهرة والى والدتي حفظها الله

إلى أغلى من وهبتهم حياتي زوجتي ... أولادي

د. محمد النعيمي

المحتويات

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

المقدمـــة

بعد الاتكال على الله سبحانه وتعالى تم الشروع في وضع الاسس والمفاهيم الاساسية لتأليف هذا الكتاب والذي كان ضرورة ملحة نظرا لما قدمه الحاسوب من تطور كبيرفي وضع الحلول الناجحة لمعالجة مهام كثيرة تتعلق في تكييف التعامل مع العلوم الصرفة في المجالين النظري والتطبيقي.

نظرا للتطور السريع في أستخدام النظم الحاسوبية ودخولها في مجالات وتطبيقات واسعة ولمواكبة هذا التطور فأنه لابد من فهمه وأدراكه بالشكل الصحيح وذلك عن طريق دراسة وتحليل كافة المجالات العلمية وتكييفها مع النظم الحاسوبية.

ونظرا لحاجة الدارسين والباحثين الى المصادر العلمية التي تغطي التطبيقات الاحصائية وتكييفها مع حزمة MATLAB تم تأليف هذا الكتاب الذي يتميز بأستخدامه عدد كبير من التطبيقات العملية الواقعية التي يحتاجها الباحث في المجال الاحصائي.

يتميز هذا الكتاب ببساطة التعامل مع المادة العلمية وأعطائها الصفة الوصفية الواضحة أضافة الى أنه قد تم التركيز بشكل وافي الى كيفية التعامل مع حزمة MATLAB والاستخدام الامثل لها في خدمة الجانب الاحصائي، حيث تفتقر المكتبة العربية الى هكذا كتب.

وما يميز حزمة MATLAB أنه يمكن التعامل معها بطريقتين: الطريقة الاولى تتمثل بالاستخدام المباشر وذلك بتطبيق الاوامر والدوال بأستخدام الشاشة الرئيسية بشكل مباشر، وأما الطريقة الثانية فتتمثل ببناء الخوارزمية اللازمة لتنفيذ كافة المهام المطلوبة بأستخدام لغة البرمجة الخاصة بالحزمة وهذا ما يميز حزمة MATLAB عن بقية الحزم الجاهزة.

وكمرحلة أولى ارتأينا أن يستعرض هذا الكتاب المبادئ الاساسية والمبسطة للربط بين حزمة MATLAB والتطبيقات الاحصائية حيث يتضمن هذا الكتاب ثمان فصول وفيما يلي وصف لها:

الفصل الاول يعرض مفاهيم عامة عن الطرق الاحصائية وكيفية عرض البيانات أضافة الى المقاييس الاحصائية.

الفصل الثاني يوضح معلومات عامة عن الحاسوب وتطوره والتقنيات والبرمجيات المستخدمة معه أضافة الى طرق بناء الخوارزميات.

الفصل الثالث يعتبر مدخل الى المفاهيم الاساسية لحزمة MATLAB حيث تم التطرق به الى كيفية تحميل الحزمة والمتكلبات اللازمة لكي يعمل النظام أضافة الى التقنيات البرمجية المستخدمة في الحزمة.

الفصل الرابع كان مخصص لتطبيقات الحزمة في مجال الاحصاء الوصفي حيث تضمن أنواع المقاييس المسخدمة في الاحصاء الوصفي مع وصف وتبويب البيانات.

الفصل الخامس يستعرض تطبيقات الحزمة في الرسومات والمخططات الأحصائية حيث تم توضيح أنواع مختلفة من المخططات والرسومات والأختبارات والتي تستخدم في أختبار العينات.

الفصل السادس يوضح تطبيقات الحزمة في أختبارات الفرضية حيث تم عرض أنوع مختلفة من الاختبارات وتطبيقاتها الأحصائية.

الفصلين السابع والثامن خصصا لتطبيقات الحزمة في النماذج الخطية والمتعددة حيث تضمنا انواع النماذج الخطية وتطبيقاتها المختلفة وأجراء كافة الاختبارات اللازمة لذلك.

وهنا لابد من الاشارة الى أن الكتاب قد تم تدعيمه بالمخططات والرسومات التوضيحية والتي لها اثر كبير في فهم وأستيعاب البيانات المعروضة.

املين من الله عز وجل أن يكون هذا الجهد المقدم بين أيدي القراء أضافة متواضعة الى المكتبة العربية التي تحتاج الى المزيد من المصادر العلمية وخاصة في مجال التعشيق والربط بين العلوم الصرفة وتطبيقاتها بأستخدام الحاسوب من حزم جاهزة وبرامجيات.

ومن الله التوفيق

المؤلفان

الفصل الأول

مفاهيم إحصائية

Fundamental Statistics

١.١ مقدمة	Introduction
1.2 الإحصاء الوصفي	Descriptive Statistics
1.3 الإحصاء الاستنتاجي	Inferential Statistics
١.٤ أساليب العينات	Methods of Samples
١.٥ طرق عرض البيانات	Methods of Presenting Data
1.6 المقاييس الاحصائية	Measurement of Statistics
1.6.1 مقاييس النزعة المركزية	Measures of central tendency
1.6.2 مقاييس التشتت والاختلاف	Measures of dispersion or variation
1.7 الارتباط	Correlation
1.8 الانحدار البسيط	Simple Regression

١.١ مقدمة Introduction

لقد وردت كلمة الإحصاء في القرآن الكريم بعدة آيات (...**وَإِن تَعُدُّواْ نِعْمَةَ اللهِ لاَ تُحْصُوهَا...**) آية (٣٤) من سورة إبراهيم، وفي سورة الكهف آية (٤٩): (...**مَا لِهَـذَا الْكِتَابِ لاَ يُغَادِرُ صَغِيرَةً وَلاَ كَبِيرَةً إِلاَّ أَحْصَاهَا...**).١

ومما تقدم نستطيع أن نتعرف على معنى الإحصاء معناه الحصر أو العد.. وقد استخدمت عمليات الإحصاء في عد أو تصنيف الأمور وتقسيمها إلى العصور القديمة في الحضارات السومرية والمصرية وحضارات الصين قبل الميلاد بعمليات جمع المعلومات أو التعداد السكاني لفرز أصناف المجتمع أو تعداد الجيوش وأصنافها... فكانت هناك عمليات عد وحصر وتوثيق هذه المعلومات في سجلات الدولة..

وأن كلمة Statistic مشتقة عن الكلمة Stat أي الدولة وهي مجموعة الحقائق الخاصة بشؤون الدولة، وعليه فإن الإحصاء له علاقة بدراسة مجتمع الدراسة أو أجزاء منه وتعني هنا بكلمة المجتمع هو مجموعة العناصر التي تخضع للدراسة ولا يقتصر الإحصاء على وصف الظواهر أو تبويبها وتصوير الأشكال التي نراها في بعض الكتب أو المجلات أو الصحف وهذه طريقة من الطرق التغييرية عن التغييرات التي حدثت للظاهرة خلال فترة زمنية معينة وأعداد البيانات وإيجاد تكراراتها حسب الفئات التي تم توزيعها.. والإحصاء هو علم يختص في جمع البيانات وعرضها وتحليلها لغرض الاستنتاج واتخاذ القرار ويساعد أيضاً متخذي القرارات بالوصول إلى القرار الأكثر صواباً من خلال توفر البيانات وطريقة جمعها.

ولو أمعنا النظر في موضوع استخدام الإحصاء لوجدنا أن هذا الموضوع يدخل في كل العلوم الأخرى فمثلاً يتم استخدام تصميم التجارب في تحليل البيانات واختيار الفرضيات وفي تحليل نماذج التنبؤ للظواهر الاقتصادية والسيطرة على المنتوج وتقدير أعداد العاملين في القطاعات المختلفة والتخطيط للقوى العاملة وفي التحليل المالي في الموازنات وأيضاً يستخدم علم الإحصاء في العلوم الاجتماعية والعلوم الإنسانية لتحليل الظواهر والحالات الاجتماعية مثل تحديد ميزانية الأسر وتحديد الجرائم وحالة الفقر

وغيرها وأيضاً للإحصاء استخدامات في مجال الصحة والدراسات السكانية في تحديد نسبة النمو والوفيات وفي مجال تحديد الأمراض ونسبة انتشار بعض الأمراض.

ومما تقدم يمكن ملاحظة أن الباحثين يحتاجون بشكل كبير إلى الأساليب الإحصائية في التحليل الكمي للوصول إلى قرارات واقعية تنطبق على الواقع الذي تم استحصال البيانات منه. إذن يلعب علم الإحصاء دوراً مهماً في كثير من القطاعات في مجمل الحياة ونشاطها وخاصة بعد التقدم التكنولوجي الكبير الذي يشهده العالم اليوم من تطور كبير في بناء نظام المعلومات والمعلوماتية ويقاس تقدم الدول على كمية المعلومات التي تحصل عليها أو توفرها إلى مجتمعاتها ولذلك سميت الثورة الجديدة بثورة المعلومات. وهذا لا يمكن الوصول له إلا من خلال الأساليب الحديثة والمتقدمة في جمع وتبويب البيانات وتحويلها إلى معلومات باستخدام أجهزة الحاسبات والبرمجيات والذي يعتبر من أهم عناصر نظم المعلومات الإدارية.

1.2 الإحصاء الوصفي Descriptive Statistics

وهو ذلك الجزء المهم من الإحصاء الذي يعني في جميع البيانات وتبويبها (تنظيمها) وتصنيفها وعرضها على شكل جداول ورسوم بيانية. إذن يستخدم الإحصاء الوصفي لوصف الحقائق وتحويلها إلى أرقام وعرضها بشكل مناسب بتطبيق المؤشرات الإحصائية التي تمثل تلك الحقائق وهناك عدة وسائل يمكن التعرف عليها من خلال ما يلي:

١. عرض البيانات Graphical Presentation: هو عرض البيانات على شكل جدول أو رسوم بيانية لتوضيح الاتجاهات العامة أو طبيعة العلاقات بين الظواهر المدروسة.

٢. المقاييس المستخدمة: وتمثل هذه المقاييس المتوسطات مثل الوسط الحسابي والوسيط والمنوال ومقاييس أخرى مثل التشتت ويمكن حسابها عن طريق حساب المدى والانحراف المعياري والتباين وهناك مقاييس العلاقات الإرتباطية وهو حساب معامل الارتباط بين المتغيرات وبيان قوة العلاقة وطبيعتها.

ويمكن تصنيف البيانات إلى بيانات وصفية (نوعية) Qualitative Data وهي المجموعة التي تصف ظاهرة معينة بصفة معينة مثل المستوى الشهادة التي يحملها الشخص مثل البكالوريوس، توجيهي، أو الحالة الاجتماعية (متزوج، أعزب) أو الجنس (ذكر، أنثى) وهكذا.

أما البيانات الرقمية (الكمية) Quantitative Data: وهي مجموعة الحقائق التي يعبر عنها بالأرقام مثل عدد السكان وعدد الموظفين في مؤسسة معينة وعدد القطع المنتجة في مصنع معين والأطوال والأوزان وغيرها من مقاييس يمكن تمثيلها بالأرقام. ويمكن تقسيم البيانات إلى بيانات منفصلة Discrete Data وهي الأرقام التي تأخذ أرقاماً صحيحة أي لا تقبل الكسور مثل عدد الأشخاص لا يمكن أن تكون مثلاً (٥.٤) خمسة أشخاص وأربعة أعشار.. لأنه غير مقبول وعليه فإن البيانات المتصلة Continuous Data فهي الأرقام التي تقبل أجزاء الرقم أي كسور الرقم مثل (٢.٣) وغيرها والأمثلة على ذلك مثل مساحة قطعة معينة (٥٠.٣) خمسون متراً وثلاثة بالعشرة من المتر وكثافة سائل معين أو زمن وصول شخص معين في الساعة (١٠.٢٠) العاشرة وعشرين دقيقة...

1.3 الإحصاء الاستنتاجي Inferential Statistics

وهي استخدام الطرق الإحصائية التي تؤدي إلى الاستنتاجات التي يتوصل إليها الباحث من خلال تحليل البيانات للوصول إلى التقديرات أو التنبؤات أو اتخاذ القرار في قبول أو رفض الفرضيات من خلال استخدام الاختبارات.

ويتضمن هذا الجانب من الإحصاء الاستدلالي أهم شيء هو التوصل إلى أفضل تقدير جيد لمعالم المجتمع:

١. التقدير Estimation

وهو تقدير معالم المجتمع من خلال سحب عينة من هذا المجتمع يوضع فترة عليا وفترة دنيا وهل هذا التقدير يقع ضمن الفترة المسموح بها أم خارجة وهذا يعني هل التقدير جيد أم لا.

٢. اختبار الفروض Test of Hypothesis

وهو إجراء اختبارات للفرضيات التي تم وضعها في بداية الدراسة بقبول الدراسة أو رفضها من خلال إجراء الاختبارات المناسبة لغرض اتخاذ القرار ونعني بهذا الجانب هو قبول الفرضية الصفرية أو الفرضية البديلة.

إذن أصبح واضحاً أن الإحصاء الاستدلالي هو معناه استخدام الطرق الإحصائية المتقدمة للتوصل إلى التقدير أو التنبؤ ومساعدة متخذ القرار بالوصل إلى قرار أكثر قرباً من الصواب.

٣. مفهوم المتغيرا Variables

يمكن تعريف المتغيرات على أساس أنها تلك الصفات أو السمات التي نقيسها على أفراد عينة معينة أو الكمية التي تتغير من عنصر إلى آخر. فمثلاً درجات الحرارة تتغير بين فترة وفترة أخرى أو أوزان مجموعة من الأطفال تتغير من وقت إلى آخر وهكذا يمكن إعطاء أمثلة كثيرة في هذا الخصوص وهذا عكس مفهوم الثابت التي تكون قيمته ثابتة رغم تغيير الزمان أو المكان مثل الكثافة النوعية لمادة معينة تكون ثابتة.

وهناك نوعين من المتغيرات:

أ- المتغيرات المنفصلة (Discrete Variables)

هو ذلك المتغير الذي يأخذ قيماً قابلة للعد أي أنها تكون محددة. ومثال ذلك عدد الطلبة الذين يدخلون الحصة المعينة فإما أن تكون (٣ أو ٤ أو ٥) وهذا يعني أن عددهم منفصل.

ب- المتغير المتصل (Continuous Variables)

هو ذلك المتغير الذي يأخذ قيم متصله وتميز منفصل وهو عبارة عن مساحة ما بين النقطتين وليس نقاط منفصلة مثال ذلك درجة الحرارة عندما ترتفع من ٢٠ درجة إلى ٢٥ درجة فإن ذلك يعني كل النقاط بين الدرجتين ومثال المساحة تحت المنحنى يتم إيجادها بطريقة التكامل بين النقطتين أي إيجاد كل النقاط.

٤. مفهوم المجتمع والعينة The concept of population and Sample

يتكون المجتمع من مجموعة عناصر أو مفردات قيد الدراسة والمفردات في المجتمع هي عبارة عن أشخاص أو عوامل أو طلاب أو أي وحدة واحدة يتكون منها المجتمع قيد الدارسة.

أما العينة هي جزء من ذلك المجتمع على أن تتمثل تلك العينة المجتمع أفضل تمثيل. أي تعميم النتائج التي تحصل عليها من خلال دراسة العينة بحيث يمكن تعميمها على مفردات المجتمع.

١.٤ أساليب العينات Methods of Samples

توجد أساليب معينة لسحب العينات وسوف نستعرض أسلوبين لسحب العينات هما:

أولا: العينات الاحتمالية (Probability Sampling) أو ما تسمى بالعينة العشوائية Random Sampling وهي فرصة ظهور كل مفردة مساوية لفرصة ظهور المفردة الأخرى وهناك عدة أساليب لسحب العينات العشوائية.

أ- العينة العشوائية البسيطة Sampling Random.

ب- العينة الطبقية العشوائية Stratified Random Sampling .

ج- العينة المنتظمة Systematic Sampling.

د- العينة العنقودية Classtiral Sampling.

ثانيا:العينات الغير احتمالية Non-Probabilistic Sampling: وأنواع هذه العينات هي:

أ- العينة الحكيمة Judgmental Sampling.

ب- العينة سهلة الاختيار Accessibility Sampling

1.5 طرق عرض البيانات Methods of Presenting Data

من المهم جداً أن يتم عرض البيانات التي تحصل عليها لغرض بيان التغيرات التي حدثت على البيانات خلال فترة زمنية أو تصنيف البيانات وفق تصنيف معين يحتاجه متخذ القرار أن يلاحظ كيف تم توزيع العينة المسحوبة أو طبيعة المشكلة التي يتعامل معها وما هي أهم التغييرات التي حدثت خلال الفترة ويعتبر هذا الأسلوب من الأساليب المهمة التي توضح لمتخذ القرار بشكل أسهل وأبسط من خلال عرض البيانات على أشكال مختلفة منها:

١- طريقة الجداول Tables

وهي وضع البيانات التي تحصل عليها بجداول مبوبة وفقاً لتصنيف معين ومثال ذلك:

حصل (١٦) طالب على علامات معينة في امتحان الفصل الأول وطلب المدرس معرفة كيفية توزيع الطلبة على مستوى العلامات.

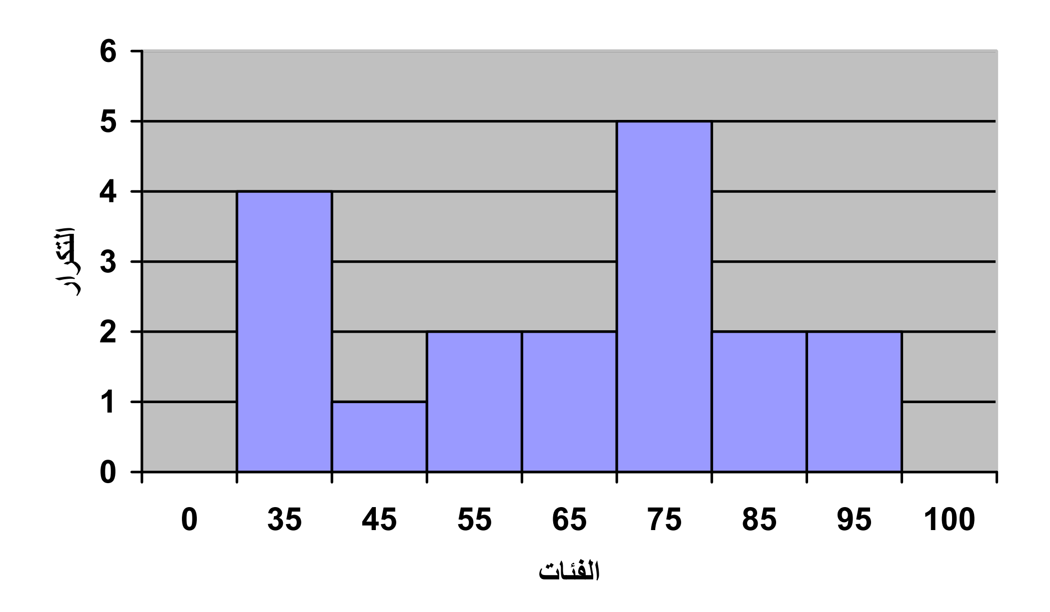
العلامات:٥٢، ٥٦، ٤٤، ٩٣، ٤٣، ٨٠، ٦٠، ٦٨، ٨٢، ٩١، ٧٢، ٧٧، ٧٩، ٨٤، ٤١، ٣٩.

الجدول:

الفئات	التكرارات بالرمز	التكرار
٣٥	////	٤
٤٥	/	١
٥٥	//	٢
٦٥	//	٢
٧٥	/////	٥
٨٥ – ٩٥	//	٢
	١٦	١٦

٢- طريقة المستطيلات Bar Graph

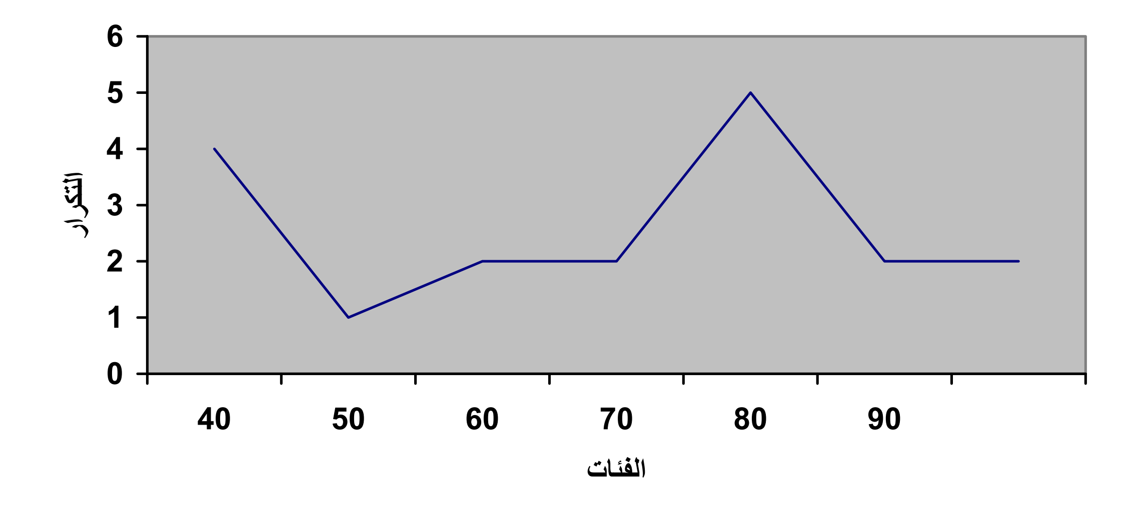
تكون هذه الطريقة بتمثيل البيانات التي يتم تبوبيها في الجدول السابق بالرسم على شكل مستطيلات حيث يكون المحور السيني للفئات والمحور الصادي العمودي للتكرارات ويمكن تمثيل البيانات على شكل مستطيلات وكما موضحة في الشكل ١.١.



شكل ١.١ رسم البيانات بطريقة المستطيلات

٣- طريقة المنحنى Curve

ويتم تمثيل هذه الطريقة من خلال إيجاد مراكز الفئات أي بقسمة حدين الفئة على (٢) أي إيجاد الوسط الحسابي بين الحد الأعلى للفئة والحد الأدنى للفئة مثال ذلك: $\frac{45+35}{2}$ وهكذا لكل الفئات ويكون الشكل وكما موضحة في الشكل ١.٢.



شكل 1.2 رسم البيانات بطريقة المنحني

٤- طريقة الدائرة Pie Chart

تتمثل هذه الطريقة بتقسيم الكل إلى أجزاء أصغر ويمثل الكل دائرة كاملة والأجزاء عبارة عن قطاعات علماً بأن قياس الزاوية الكلية للدائرة ٣٦٠° (أي نسبة القطاع مضروبة في الزاوية الكلية).

المثال السابق تمثيل علامات الطلبة كما موضحة في الشكل ١.٣ حسب أعدادهم يكون:

عدد الطلبة (١٦)

الفئة الأولى $$٩٠^{\circ} = \frac{4}{16} \times ٣٦٠^{\circ}$$

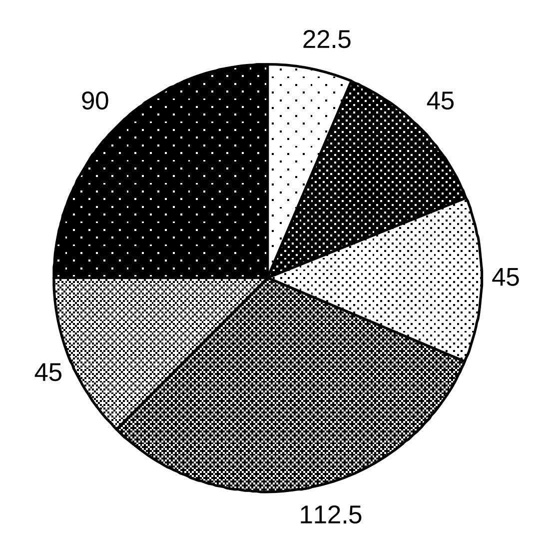
الفئة الثانية $$٢٢.٥^{\circ} = \frac{1}{16} \times ٣٦٠^{\circ}$$

الفئة الثالثة $\text{٣٦٠}^{\circ} \times \frac{2}{16} = \text{٤٥}^{\circ}$

الفئة الرابعة $\text{٣٦٠}^{\circ} \times \frac{2}{16} = \text{٤٥}^{\circ}$

الفئة الخامسة $\text{٣٦٠}^{\circ} \times \frac{5}{16} = \text{١١٢.٥}^{\circ}$

الفئة السادسة $\text{٣٦٠}^{\circ} \times \frac{2}{16} = \text{٤٥}^{\circ}$



شكل ١.٣ رسم البيانات بطريقة الدائرة

1.6 المقاييس الإحصائية Measures of Statistics

سنتناول في هذه الفقرة أهم المقاييس التي ستستخدم في قياس بعض من الظواهر وإيجاد مؤشرات لها من خلال وصف الظاهرة وإعطاء مؤشرات تبين طبيعة الظاهرة وكيفية توزيع البيانات الموجودة فيها من خلال قياس الأوساط الحسابية وتحسن مقاييس النزعة المركزية ومقاييس التشتت والتعرف على الانحراف المعياري ومعامل الاختلاف وغيرها ويستفاد من هذه المقاييس متخذ القرار للتعرف على طبيعة البيانات ويمكن أيضاً الاستفادة منها في المقارنة بين العينات واختيار الأفضل.

1.6.1 مقاييس النزعة المركزية Measures of central tendency

في كثير من التوزيعات التكرارية نجد أن عدد كبير من المفردات يميل إلى التجمع حول قيمة متوسطة معينة ويقل عدد المفردات تدريجياً كلما بعدنا عن القيمة المتوسطة التي تمثل مركز التوزيع وتسمى هذه الظاهرة بالنزعة المركزية وهي نزوع المفردات إلى التجمع حول مركز التوزيع. وهذا يعتمد على طبيعة البيانات وتكراراتها ولذلك فإن لكل مجموعة بيانات وسط حسابي يختلف من المجموعة الأخرى والتي يمكن استخدام ذلك لوصف المجموعة التي تحدد مركز أو متوسط المجموعة.

إذا أراد مدير أن يعرف متوسط إنتاجية مجاميع من العاملين بعد أن يقوم بتقسيمهم إلى مجاميع حسب المواصفات أو سنوات الخدمة أو الخبرة وغيرها من أمور يمكن أن يمتد متوسط إنتاجية كل مجموعة والتي على أساسها نقيس إنتاجية الأفراد ويمكن أيضاً أن يتخذ قرارات بهذا الصدد لكل مجموعة وفقاً للمعايير التي تم اعتمادها. وهناك عدة معايير يمكن اعتمادها لغرض التعرف على ما ذكر في إيجاد الفروق بين المجموعات:

١. الوسط الحسابي Arithmetic Mean

٢. الوسط Median

٣. المنوال Mode

٤. الوسط الهندسي geometric mean

وأن اختيار أحد هذه المقاييس يتوقف على طبيعة البيانات التي ندرسها كما يتوقف على الهدف الذي ننشده.

وتجدر الإشارة لهذه المقاييس مميزات وعيوب لكل واحدة منها ويمكن تلخيصها بما يلي:

أ- يجب استخدام كافة المفردات في حال حساب مقاييس النزعة المركزية لكي تكون ممثلة للبيانات أفضل تمثيل.

ب- يفضل أن تكون القيمة المتوسطة من القيم التي يكون تأثيرها بذبذبات العينة قليلاً فإذا كان عدد العينات المأخوذة من مجتمع واحد فمن النادر أن تتساوى متوسطات هذه العينات مهما كانت هذه المتوسطات ولكن من المحتمل أن تكون المتوسطات قريبة بعضها عن البعض.

ج- يفضل أن تكون القيمة المتوسطة قيمة موضوعية ليست قيم متعددة ذاتياً من الباحث ويفضل أن تحسب بالطرق الرياضية.

وتعتبر طرق مقاييس النزعة المركزية من الطرق المهمة جداً في الاستخدامات الإحصائية وتفيد الباحثين ومتخذي القرارات بشكل كبير وتعتبر سهلة الاستخدام.

١.٦.٢ مقاييس التشتت والاختلاف

Measures of dispersion or variation

يقصد بالتشتت هو درجة التفاوت أو الاختلاف بين قيم هذه المجموعة فمثلاً إذا كانت قيم المجموعة قريب بعضها عن بعض يكون التشتت صغيراً وإذا كانت الحالة عكس ذلك أي القيم متباعدة يكون التشتت عالي وكبير. ولذلك يعتبر مقاييس التشتت مقياساً للتجانس قيم المجاميع أي كلما قل التشتت كان التجانس كبير والعكس إذا كانت قيم المجاميع متباعدة كان التشتت كبير. ونعني بالتجانس هو اقتراب القيم من وسطها الحسابي. ومن أهم هذه المقاييس التي تستخدم هي:

١. المدى Range

وهو الفرق بين أكبر قيمة في المجموعة والقيمة الصغرى.

٢. التباين Varianceوهو من أكثر المقاييس استخداماً وهو عبارة عن قسمة مجموع مربعات انحرافات القيم عن وسطها الحسابي على عددها وذلك للتمكن من التغلب على مشكلة الإشارات عند جمع الانحرافات التي تؤدي ظهور الإشارة السالبة والموجبة والتي تؤدي إلى أن مجموع الانحرافات تساوي صفراً وعليه تم تربيع الفرق بينهما.

٣. الانحراف المعياري Standard Deviation

وهو عبارة عن الجذر التربيعي للتباين .

٤. الدرجة المعيارية Standard Scores

وتستخدم الدرجة المعيارية للمقارنة بين مفردتين في مجموعتين مختلفتين ويجب تحويل وحدات كل مفردة إلى وحدات قياسية حتى تكون المقارنة صحيحة أي أبعاد وحدات القياس لكل مفردة من بعد تحويلها إلى قيم معيارية ويرمز لها (Z) ونحصل عليها من خلال إيجاد الفرق بين القيمة الأولى ووسطها الحسابي لتلك المجموعة ثم نقسم الفرق على الانحراف المعياري للمجموعة الأولى.

٥. معامل الاختلاف Coefficient of variation

وهي إجراء مقارنة بين مجموعتين من القيم أو التوزيعين من حيث تشتتها أو اختلافها ويرمز له بالحرف (CV) والذي يمكن حسابه من خلال قسمة الانحراف المعياري على الوسط الحسابي للمجموعة. ويمكن مقارنتها مع المجموعة الثانية.

1.7 الارتباط Correlation

وهو العلاقة بين ظاهرتين ويمكن قياس هذه العلاقة من خلال إيجاد معامل الارتباط الذي يكشف طبيعة العلاقة هل هي علاقة قوية أم ضعيفة علاقة إيجابية أم علاقة سلبية ولذلك فإن معامل الارتباط الذي يرمز له بالحرف (r) تكون حدوده.

$-1 \leq r \leq +1$

وهناك عدة أنواع من العلاقات منها:

١. علاقة سببية Causal Relationship

وهو إذا كان التغير في إحدى الظاهرتين حصل بسبب التغيير في الظاهرة الأخرى فإن العلاقة بينهما تسمى علاقة سببية مثال العلاقة بين السعر والكمية من بضاعة معينة.

٢. العلاقة الغير مباشرة Indirect Relationship

وهو حصول تغير في الظاهرتين بسبب تغير في الظاهرة الأخرى بسبب غير مباشر فمثلاً زيادة الرسوم الكمر كية على سلع معينة يسبب زيادة أجور العمل لأن زيادة الرسوم الكمر كية تسبب زيادة سعر السلع وهذا يؤثر على زيادة الشراء على السلع المماثلة للسلع المحلية وهذا يؤدي إلى زيادة إنتاجها وهذا يسبب زيادة الطاقة الإنتاجية وزيادة ساعات العمل للعمال مما يسبب زيادة الأجور.

٣- علاقات عرضية

أو ما تسمى بالعلاقات المصادفة وهو تأثير ظاهرتين نتيجة وجود عامل واحد يؤثر على الظاهرتين ويصبح كل تغير في أحداهما مرافقاً للتغير في الأخرى فمثلاً العلاقة بين سعري بضاعتين يتم شرائهما من قبل شريحة معينة من المجتمع فزيادة القوة الشرائية لهذه الشريحة يؤدي إلى زيادة سعر هذه البضاعتين معاً وتكون العلاقة طردية أو عكسية حسب طبيعة العلاقة بينهما. وهناك أنواع من المقاييس الارتباطية منها:

- **الارتباط البسيط (Simple correlation Coefficient)**

وهو قياس درجة الارتباط وقوته بين الظواهر فإذا كانت العلاقة بين الظاهرتين خطية تسمى العلاقة لإيجاد معامل الارتباط بالارتباط البسيط.

- **الارتباط المتعدد Coefficient of multiple - correlation**

وهو عبارة عن تغيير في ظاهرة معينة ينتج عنه تغيير في ظواهر أخرى ويراد قياس مقدار العلاقة بين تلك الظاهرة بكل الظواهر مجتمعة. أي إذا كان لدينا المتغيرات x_1, x_2, x_3

وأردنا معرفة العلاقة بين (x_1, x_2) و (x_1, x_3) و (x_1, x_2, x_3) فإن مثل هذه العلاقة يمكن أن نجد معامل الارتباط المتعدد وفق صيغ رياضية معتمدة لذلك.

ويمكن التفسير عنها بالرموز (r_{12}) أي العلاقة بين x_1, x_2 (r_{23}) العلاقة بين $(x2, x3)$ و $(r_1.23)$ العلاقة بين x2, x3, x1 وهكذا. وحسب عدد الظواهر المراد قياس معاملات الارتباط وهناك عدة طرق لحساب معاملات الارتباط يمكن أن يستخدمها الباحث ومتخذ القرار حسب طبيعة الظاهرة وطبيعة البيانات المستخدمة في الظواهر. فمثلاً هناك معامل الارتباط الرتبي (rank Correlation) ، وهناك معامل الارتباط النوعي للظواهر التي تكون مفرداتها تعتمد على التصنيف النوعي مثل ذكر وأنثى أو الدرجات العلمية وغيرها ويسمى معامل الارتباط (ϕ) وهناك معامل الارتباط المتسلسل والذي يسمى بأي سريال . وهكذا من معاملات الارتباط التي يمكن الاطلاع عليها والتعرف عليها بتعمق من خلال مراجعة كتب ومراجع الإحصاء.

أما معامل الاقتران والذي تحتاجه في معرفة درجة العلاقة بين ظاهرتين لا يمكن التعبير عنها بالأرقام مثل الارتباط بين التطعيم بمصل معين لمرض والإصابة بهذا المرض أو الحالة التعليمية لمجموعة من الأشخاص ومستواهم المعاشي والتي لا يمكن حساب الأوساط الحسابية في ظواهر غير رقمية.

وهناك معامل آخر يسمى معامل التوافق. والذي يستخدم في الظواهرالتي تنقسم إلى نوعين. مثل مجموعةمن الطلبةتنقسم علاماتهم إلى مستويات مختلفةمثل جيدجداًوجيدومقبول في مساقين علميين وأردنامعرفة الطلبةالذين يحصلون تقديرمعين للمساقين فيجب أن تأخذجدول يضم بعدين الأول تقديرالمساق الأول والثاني والبعدالثاني للمساق الآخرونجدمن خلال الجدول عددالطلبةالذين حصلوا على كل تقديرفي مساقين، وتعتمدهذه الطرق لإيجادمعاملات الاقتران ومعامل التوافق لحالات التجارب العلميةلكثيرمن الحالات وخاصةفي الجانب الإداري لمعرفةالعلاقة بين الظاهرتين وفق صيغ معينةويحتاجهامتخذالقرارلكي يتم عرضهاوفق جداول معينة لغرض تحديد نوع العلاقةوالنسب التأثيرية لكل ظاهرة على الظاهرةالأخرى داخل المجموعةالواحدة التي تضم أكثر

من متغير، وأن كل ما تقدم من شرح للمقاييس الإحصائية في مجال النزعة المركزية أو مقاييس التشتت أو الارتباط بكافة أنواعه يعتبر من أنواع الإحصاء الوصفي لكونه يصف البيانات.

1.8 الانحدار البسيط Simple Regression

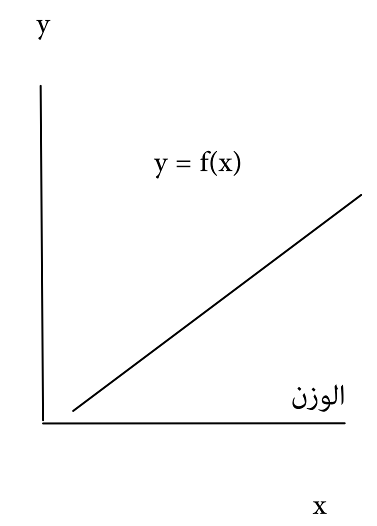
الانحدار البسيط هو معادلة خطية تربط بين ظاهرتين الأولى تسمى بالمتغيرات المعتمدة والمتغيرات المستقلة والعلاقة بينهما هي علاقة سببية أي عندما يتأثر المتغير المستقل يؤثر على المتغير المعتمد في زيادته أو نقصانه والعلاقة بينهما يجب أن تكون علاقة سببية من خلال معادلة الانحدار (Regression equation) وإذا رمزنا للمتغير المستقل مثلاً (x) والمتغير المعتمد (y) فإن المعادلة الخطية التي تجمعها هي:

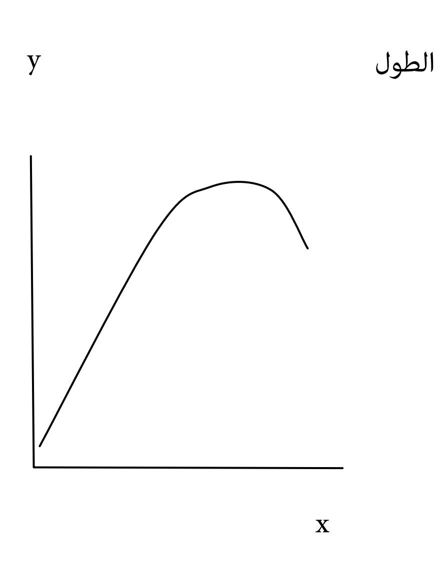
$$(y = \alpha + \beta x)$$

وتسمى هذه المعادلة بمعادلة انحدار (y) على (x) أي أننا نسعى للحصول على تقديرات معادلة يمكن استخدامها في إيجاد مؤشرات لتقدير المتغير المعتمد (y) أي يمكن استخدام هذه المعادلة لأغراض التنبؤ بقيم (y).

وتبدأ عملية تقدير معادلة الانحدار بجمع البيانات التي تمثل أزواج المتغيرين المتناظرة. ويعدها تقرير نوع العلاقة بينهما هل هي علاقة خطية أو غير ذلك من خلال المعرفة المسبقة والخبرة هل العلاقة خطية أم لا؟ أي علاقة أسية مثلا يعني زيادة (x) تسبب في زيادة (y) لفترة معينة، وبعدها يظهر التأثير سلبي أي نقصان الظاهرة (y) يكون مثل تلك الحالات معادلة أسية والأمثلة كثيرة يمكن التعرف عليها.

حالة العلاقة الخطية مثلاً طول الشخص ووزنه فكلما زاد طول الشخص فأن وزنه يزداد. أما الحالة الأخرى فإن تناول مادة معينة تكون فائدتها جيدة وفي حال زيادة تناول نفس المادة في نفس الوقت يكون الأثر سلبي عليها ، ويمكن تمثيل العلاقة بينهما بالشكل ١.٤.





زيادة (x) تؤدي إلى زيادة (y) دائما   زيادة x تؤدي إلى انخفاض (y) بعد فترة

شكل ١.٤ رسم العلاقة الخطية واللاخطية

ولتقدير معادلة الانحدار يتم استخدام طريقة المربعات الصغرى (Ls) التي من خلالها يمكن تقدير معاملات النموذج الخطي الذي تم التنويه عنه ولكن سوف يتم كتابة معادلة الانحدار وتوضيح كل مفرداتها كما يلي : $y = a + b x + e$
حيث أن :

y : المتغير المعتمد (dependent variable)

x : المتغير المستقل Independent variable

a : الحد الثابت أو معامل التقاطع

b: مقدار ما تضيف وحدة واحدة من المتغير المستقل إلى قيمة (y) وترتبط إشارة (b) باتجاه الارتباط بين (xy) أي عندما تكون الإشارة موجبة تكون العلاقة موجبة والعكس.

e : حد الخطأ أي هو الفرق بين القيم الأصلية (y) والقيمة التقديرية (y).

ويمكن تطوير نموذج الانحدار البسيط إلى الانحدار المتعدد أي بإضافة متغيرات مستقلة أخرى للنموذج ويمكن حساب المعلمات (Parameters) من خلال المعادلات الرياضية وطريقة المصفوفات (Mattresses) لغرض التوصل إلى قيمة هذه المعلمات ويمكن أيضاً التنبؤ بالظاهرة من خلال التعويض عن قيم المعلمات والمتغيرات المستقلة للوصول إلى تقدير أو التنبؤ بقيمة ($\hat{y}$) وإيجاد الخطأ المتوقع من خلال فحص النموذج. وتعتبر العلاقة بينهما علاقة خطية وأيضاً هناك نماذج متعددة المتغيرات الغير خطية يمكن بيانها بما يلي:

١- نموذج خطي متعدد:

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

٢- نموذج غير خطي متعدد :

$$y = a + b_1 x_1 + b_2 x_{2^2} + b_3 x_{\backslash 3^3} \dots + b_n x_n$$

وتحتاج عملية المعادلات المتعددة الخطية وغير الخطية لاستخدام البرمجيات في الحسابات لغرض التوصل إلى إيجاد المعلمات ويمكن فحص جودة التقدير من خلال البيانات المتوفر وجودة التقدير ويمكن استخدام ذلك من خلال فحص المصطلحات باختبار (F) في جدول تحليل التباين والذي سيتم التطرق له في الفقرات القادمة من خلال اختبار الفرضيات واختبارها.

الفصل الثاني

الحاسوب والمعلوماتية

Computer and Information

2.1	مقدمة	Introduction
2.2	انواع البيانات	Data Types
٢.٣	مكونات الحاسوب	Components of Computer
٢.٤	خصائص الحاسوب واستخداماته	Computer Performances and Applications
٢.٥	اجيال الحواسيب	Computers Generations
٢.٦	الخوارزميات	Algorithms
٢.٧	حزمة MATLAB	MATLAB Package

٢.١ مقدمة Introduction

الإنسان ومنذ العصور القديمة كان يميل إلى الترتيب والتوثيق وعلى هذا الأساس نشأت أولى الحضارات القديمة في وادي الرافدين والنيل. لقد كانت الرسومات من طرق التوثيق القديمة حيث كانوا يدونون الإعمال والأحداث على شكل صور ورسوم ذات دلالة. بعد ذلك ظهرت الكتابات الرمزية كمرحلة متقدمة عن الرسومات حيث كانوا يعبرون عن الأشياء برموز متفق عليها وتطور ذلك إلى ظهور الكتابة المسمارية في وادي الرافدين.

بتطور وتوسع الإعمال التجارية والزراعية والصناعية أصبح التعامل مع البيانات والمعلومات بشكل كبير جدا حيث اصبحت الحاجة ملحة لتجاوز الاعمال والحسابات اليدوية الى مكننة العمل. اذاً فالسبب الرئيسي وراء مكننة العمل وتطورها وإيجاد الحواسيب هو الحاجة إلى معالجة كم الكبير من البيانات سواءاً من الناحية الرياضية او الإحصائية لتأدية او تنفيذ المهام أو الأعمال الكبيرة. أولى الحواسيب التي صممت وأستخدمت لاجراء وتنفيذ العمليات والمهام كانت هي الحواسيب اليدوية ثم تلتها الحواسيب المكانيكية وبعدها جاءت الحواسيب الإلكترونية والرقمية ومن أمثلتها المتداولة والفعالة هي الحواسيب الشخصية وظهرت كذلك الحواسيب فائقة السرعة وذات التطبيقات الخاصة غيرها .

وقد دخل الحاسوب الى العالم بشكل تجاري في منتصف الخمسينيات حيث حدثت طفرة كبيرة في عالم الإلكترونات وعالم الأقتصاد . وبتطور الحاسوب تطورت معها أجهزة الخزن أبتداءا من الأشرطة الممغنطة إلى الأقراص المرنة والأقراص الصلبة والأقراص الليزرية والذاكرة المتنقلة وغيرها حيث ازدادت قدرتها على خزن البيانات بشكل كبير جدا حيث أصبح بالإمكان الآن تخزين مجلدات وكتب كاملة على قرص ليزري صغير. إضافة إلى القدرة الخز نية العالية لأجهزة خزن المعلومات الحديثة فأنها امتازت بمرونتها وسهولة التعامل معها ونقل المعلومات منها وأليها وكذلك سهولة تخزينها وحملها.

ظهور الحواسيب الشخصية في مطلع الثمانينيات ادت الى حدوث طفرة كبيرة في عالم الإلكترونيات حيث سهلت امور كثيرة في معالجة وادخال واخراج البيانات . وشهدت اجهزة الحاسوب تطورا كبيرا من حيث السرعة والخزن وبالنسبة للسرعة ازدادت من عشرات

الميغاهرتز الى اكثر من اربعة غيغاهرتز واما الخزن فازداد من عشرات الميغابايت الى اكثر من مائتي غيغابايت . ان هذا التطور الكبير في الحواسيب كأجهزة رافقة تطور كبير في برمجيات الحاسوب من انظمة تشغيل وتطبيقات فظهرت انظمة التشغيل متعددة المهام ولغات البرمجة الحديثة والتقنيات ذات الوسائط التعددة .

اضافة الى ذلك شهد العالم طفرة كبيرة في عالم الإتصالات وتكنولوجيا المعلومات حيث أصبح هذا الحقل هو مركز استقطاب للإقتصاد ورأس المال وذلك بإنشاء مشاريع ضخمة في هذا المجال. يعتبر الانترنت والاتصالات المتنقلة والخلويات من أهم تقنيات هذا العصر حيث اصبحت تقدم تقنيات هائلة في نقل وتبادل المعلومات عبر كل بقاع العالم. العالم اليوم أصبح كأنه قرية صغيرة فبإمكانك أن تحصل على أي معلومة وأن تعمل أي صفقة تجارية وأن تشارك في أي مناقشة وأنت موجود في مكتبك كل هذا فتح أبواب واسعة لتبادل المعلومات والاستفادة من ذلك في تطبيقات ومجالات مختلفة.

٢.٢ أنواع البيانات Data Types

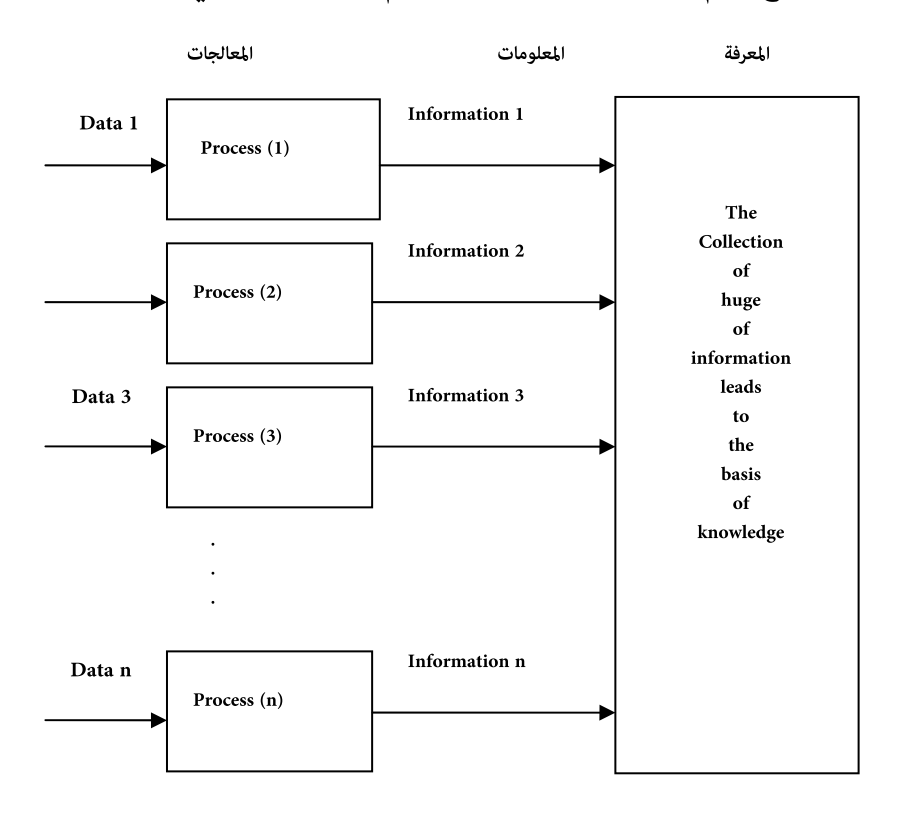
البيانات والتعامل معها ومعالجتها يرجع الى التاريخ القديم حيث ظهرت الحاجة الى ادارة وترتيب هذه البيانات خصوصا بعد تطور الأعمال التجارية والإقتصادية. لقد أستخدم الانسان على مرور الزمن وحسب فترات التطور الحضاري أدوات مختلفة لتوثيق ومعالجة البيانات فمنها اليدوية ومنها الميكانيكية ومنها الالكترونية. وكان لظهور الحاسوب أثر كبير في تنظيم البيانات وأجراء المعالجات اللازمة عليها وكان له دور كبير في توظيف هذه البيانات وخزنها ومعالجتها وتحليلها وأستخدامها في تطبيقات مختلفة. ويمكن تصنيف البيانات من حيث الاستفادة منها كما موضحة في الشكل ٢.١ الى:

١- البيانات وهي المواد الخام الإدارية وتكون عادة مبهمة او غير مفهومة بالنسبة للمقابل لذا تحتاج الى معالجة او تحويلها من شكلها الحالي الى هيئة معلومات ذات قيمة يمكن الاستفادة منها .

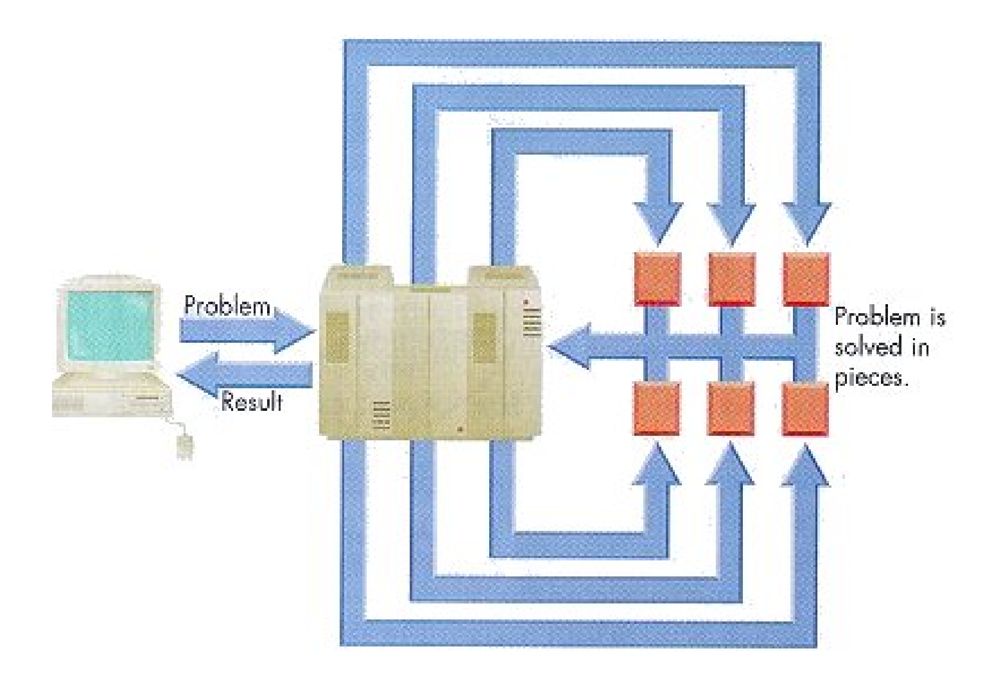
٢- المعلومات وهي المواد الناتجة عن اجراء معالجة معينة على البيانات لكي تكون على هيئة او شكل منسق مفهوم حيث يمكن للإنسان الإستفادة منها في تأدية مهمة أو عمل معين .

٣- المعرفة وتمثل النتيجة النهائية للمعلومات حيث من المعلومات يمكن التوصل الى معرفة معينة بإتجاه معين وهي شيء جديد يضاف الى معلوماتنا ، وان مجموعة المعلومات في حقل معين تمثل المعرفة في ذلك الحقل .

أن ما يدور في جهاز الحاسوب من معالجات للبيانات يتضمن أدخالها الى وحدة المعالجة المركزية عن طريق وحدة أدخال البيانات ومن ثم تخرج المعلومات الى وحدة الاخراج ليتم مراقبتها من قبل المستخدم وكما موضحة في الشكل ٢.٢.



شكل ٢.1 مخطط معالجة البيانات



شكل ٢.٢ كيف تقوم وحدة المعالجة المركزية بحل المهام

٢.٣ مكونات الحاسوب Components of Computer

الحاسوب يستقبل البيانات والمعلومات عندما يكون على الشكل الرقمي أي صفر و واحد اي ان كل المعلومات يجب أن تحول الى هذا الشكل لكي تكون مقبولة من قبل الحاسوب . البت (Bit) يعتبر اصغر وحدة لتمثيل البيانات وكل ثمان بتات تمثل بايت (Byte) ويمكن بهذا البايت تمثيل كافة الرموز والأرقام والحروف وادخالها للحاسوب. عمليات المعالجة وأتصال قطع الحاسوب مع بعضها ومع المعالج المركزي يمكن ملاحظته بوضوح في الشكل ٢.٣ . أما بالنسبة الى الوحدات والمكونات الاساسية للحاسوب كما موضحة في الشكل ٢.٤ فيمكن تقسيمها الى ما يلي:

١- وحدة المعالجة المركزية : وتقوم بمعالجة البيانات والسيطرة على جميع أجزاء الحاسوب الاخرى .

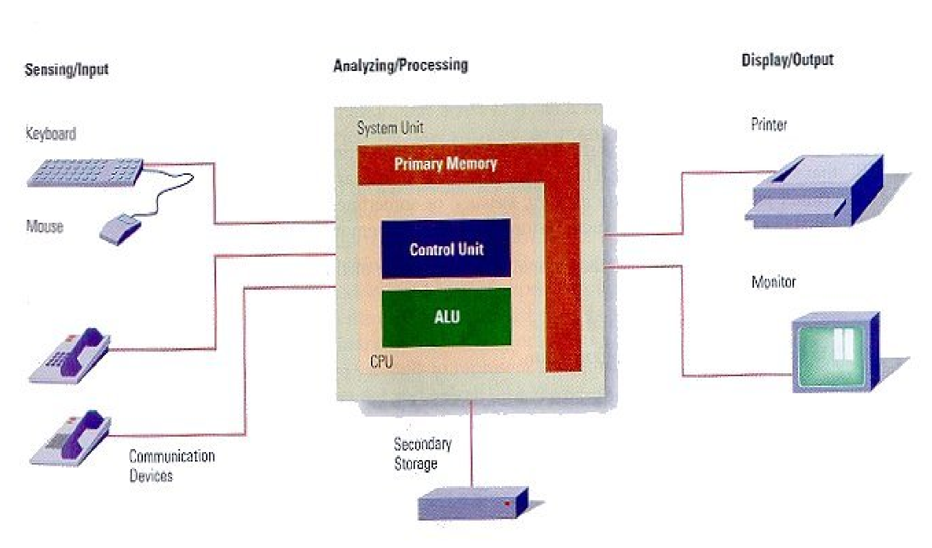
٢- وحدة الخزن الرئيسية : وهي الذاكرة الرئيسية وتقوم بخزن البيانات والبرامج بشكل مؤقت خلال عملية المعالجة .

٣- وحدة الخزن الثانوية : وتقوم بخزن البيانات والبرامج بشكل دائمي عندما لا تكون مهيئة للمعالجة مثل الأقراص الممغنطة والمرنة والليزرية .

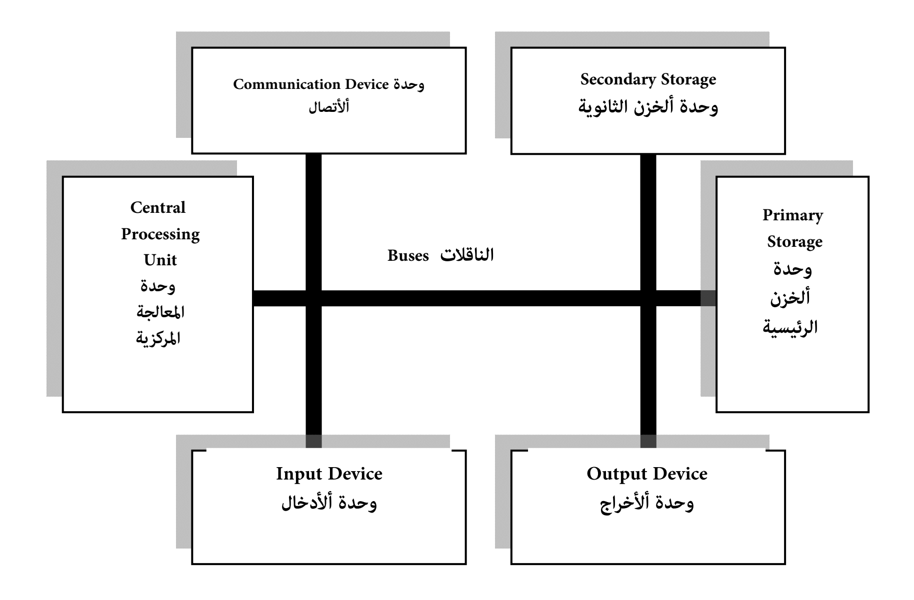
٤- وحدة الادخال : وتقوم بإدخال البيانات والمعلومات للحاسوب من أجل المعالجة وتشمل لوحة المفاتيح وشاشة اللمس والفأرة و الميكروفون والكاميرا وغيرها.

٥- وحدة الإخراج : وتقوم بإخراج وعرض البيانات والمعلومات لكي يراه المستخدم وتشمل شاشة الحاسوب والطابعات بإنواعها وغيرها.

٦- وحدة الإتصال : وتقوم بالسيطرة على مرور وانتقال المعلومات والبيانات بين كافة أجزاء الحاسوب وهو بعتبر الناقل والوسيلة لنقل البيانات والمعلومات .



شكل ٢.٣ كيفية الاتصال بين وحدات جهاز الحاسوب



شكل ٢.٤ مكونات الحاسوب

2.4 خصائص الحاسوب واستخداماته

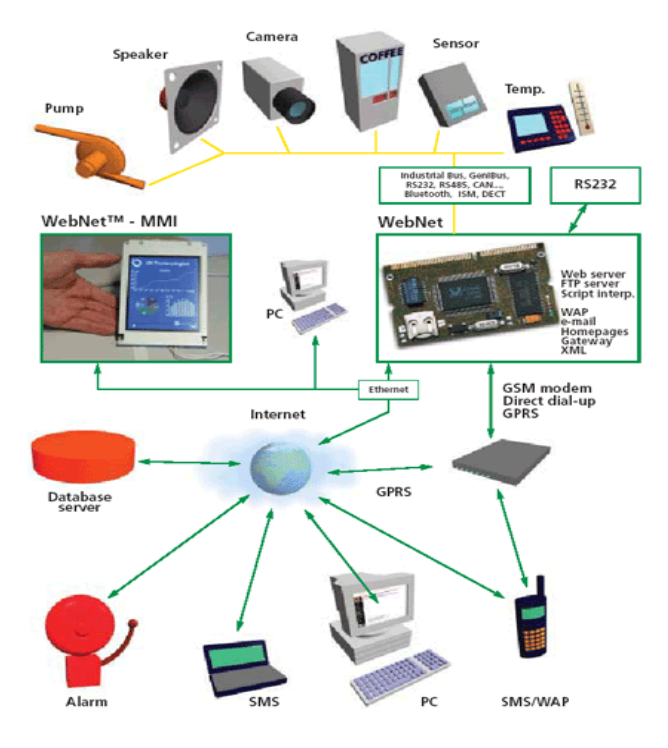
Computer Performances and Applications

الحاسوب جهاز الكتروني يقوم بادخال البيانات عن طريق اجهزة الادخال ومعالجة هذه البيانات في المعالج المركزي وتخزين البيانات في أجهزة الخزن ثم تعرض بعد ذلك بإرسالها إلى أجهزة الاخراج. ولقد مر الحاسوب بتطورات كبيرة منذ أن بدأ كجهاز بسيط إلى أن أصبح جهاز معقد يدير كافة المهام والأعمال. ومن الخصائص المميزة لهذا الجهاز هو أنه يعتمد على دقة التعامل مع البيانات وسرعة معالجة البيانات والطاقة التخزينية للبيانات وكذلك قدرته غلى الاتصال والتعامل مع الاجهزة الاخرى.

للحواسيب استخدامات واسعة جدا في عالمنا اليوم ويمكن تصنيفها الى المجالات الصناعية والمجالات التجارية والبورصةوالاسواق والمجالات الخدمية والدعائية والمجالات العلميةوالتعليميةودراسة حالةالطقس والتنبؤات الجويةوالمجالات الطبية والتشخيصية

والمجالات العسكرية والمراقبة والتجسس وأستراق المعلومات ومجالات الاتصالات والأقمار الاصطناعية، إضافة ذلك فأنه شهد تقدما كبيرا في عالم الإيصالات والمعلوماتية والتطور السريع لهذا العالم ودخول شبكات الحاسوب بأنواعها المختلفة وانتشارها الواسع كل ذلك قد ساعد في حدوث طفرة هائلة في عالم الاقتصاد والسوق والاستثمار.

الحاسوب اليوم أصبح أداة فعالة وضرورية في أداء وتنفيذ مهام كثيرة جدا ويمكن القول أنه أصبح من مستلزمات الحياة الضرورية وأن الاستغناء عنه بدا أن يكون غير ممكن وغير معقول وأصبح مقياس تطور البلد يقاس بعرفتهم وأستخدام للحاسوب والتقنيات التكنولوجية المرفقة له ومنها الانترنيت وشبكات الحاسوب والاتصالات بأنواعها المختلفة كالأقمار الاصطناعية والاتصالات المتنقلة وغيرها. فباستخدام الحاسوب وتقنياته الحديثة بأمكانك أن تدير وتسيطر على كافة الاعمال والاجهزة في البيت أو العمل بشكل أني عن طريق جهاز الخلوي أو أي طريقة حديثة للاتصال كما موضحة في الشكل ٢.٥.



شكل ٢.٥ أستخدامات وتطبيقات الحاسوب

٢.٥ أجيال الحواسيب Computers Generations

منذ ظهور الحواسيب الالكترونية في بداية الخمسينيات حدث لها تطورات كثيرة مثل تخزين البيانات فظهرت اجيال متعددة من الحواسيب فيها:

الجيل الاول: ظهر هذا الجيل في بداية الخمسينيات وتعتمد على الصمامات في العمل تصل سرعتهاالى ٢٠ الف عملية في الثانية وكانت تعتمد في تنفيذ العمليات على لغة الماكنة في كتابة البرامج .

الجيل الثاني : ظهر هذا الجيل من عام ١٩٥٩ الى ١٩٦٥ ويعتمد على الترانزيستور في العمل ، وسرعتها تصل في تنفيذ العمليات الى مئات الالاف عملية في الثانية الوحدة ، واستخدمت لغات برمجة راقية مثل كربول وفورتران .

الجيل الثالث : ظهر هذا الجيل بين ١٩٦٥-١٩٧٠ ويعتمد على الدوائر المتكاملة والمصنوعة من رقائق السيلكون وسرعتها وصلت الى نانو ثانية واستخدمت معها اجهزة ادخال واخراج سريعة.

أجيال الحواسيب الشخصية وهي:

الجيل الاول للحواسيب الشخصية : انتجت شركة انتل Intel المعالج ٨٠٨٦ في عام ١٩٧٨ والذي حقق طفرة بالنسبة للحواسيب الشخصية ثم تلى ذلك المعالج ٨٠٨٨ في عام ١٩٧٩ . انتجت شركة IBM الحاسوب الشخصي عام ١٩٨٢ والذي يعتمد على المعالج ٨٠٨٨ وكانت سرعة ذاكرته ٦ كيلو بايت وسعته التخزينية ٣٠ ميكابايت. وبعدها انتجت IBM الحاسوب الشخصي XT في عام ١٩٨٤ حيث كانت سعة ذاكرته ٦٤ كيلوبايت.

الجيل الثاني للحواسيب الشخصية : انتجت شركة JBM جهاز الحاسوب AT في عام ١٩٨٦ والذي يعني ذر التقنية المتطورة ويعتمد على المعالج ٨٠٢٨٦ وكانت ذاكرته ٤ ميكابايت وسعته التخزينية ٧٠ ميكابايت وسرعته ١٦ ميكاهرتز .

الجيل الثالث للحواسيب الشخصية : انتج هذا الجيل من الحواسيب الشخصية عام ١٩٨٨ ويعتمد على المعالج ٨٠٣٨٦ وكانت ذاكرته ١٢٨ ميكابايت ةسغته التخزينية ٦٠٠ ميكابايت وسرعته ٣٣ ميكاهرتز وتوجد معه تقنيه تخزينية مؤقتة .

الجيل الرابع للحواسيب الشخصية : انتج هذا الجيل من الحواسيب الشخصية في عام ١٩٩٠ (بداية التسعينيات) ويعتمد على المعالج ٨٠٤٨٦ وكانت ذاكرته ٢٥٦ ميكابايت وسرعتة ٦٦ ميكاهرتز .

الجيل الخامس للحواسيب الشخصية : ظهرت أجيال الحاسبات الشخصية التي تحمل المعالجات بنتيوم في منتصف التسعينات ويمتاز بأنه ذو تقنية فائقة حيث بإمكانه تنفيذ اكثر من أمر في آن واحد ومن أنواع هذه المعالجات :

١- بنتيوم 1 كانت سرعته ٧٥ ميكاهيرتز ثم تطورت إلى ٩٠ ميكاهيرتز و إلى ١٣٣ ميكاهيرتز والى ١٦٦ ميكاهيرتز ثم إلى ٢٠٠ ميكاهيرتز وكانت سعته التخزينية ١.٢ كيكابايت .

٢- بنتيوم 2 كانت سرعته ٢٣٣ ميكاهيرتز ثم تطورت إلى ٢٦٦ ميكاهيرتز ثم إلى ٣٣٣ ميكاهيرتز ثم إلى ٤٠٠ ميكاهيرتز و إلى ٤٥٠ ميكاهيرتز و ٥٠٠ ميكاهيرتز و أصبحت سعته التخزينية ٦ كيكابايت .

٣- بنتيوم 3 كانت سرعته ٦٠٠ ميكاهيرتز ثم تطورت إلى ٨٠٠ ميكاهيرتز ثم إلى ٨٣٣ ميكاهيرتز و إلى ١ كيكاهيرتز و أصبحت سعته التخزينية ٣٠ كيكابايت .

٤- بنتيوم 4 كانت سرعته ١.٢ كيكاهيرتز ثم تطورت إلى ١.٥ كيكاهيرتز و إلى ١.٧ كيكاهيرتز و أصبحت سعته التخزينية إلى ٤٠ كيكابايت .

٢.٦ الخوارزميات Algorithms

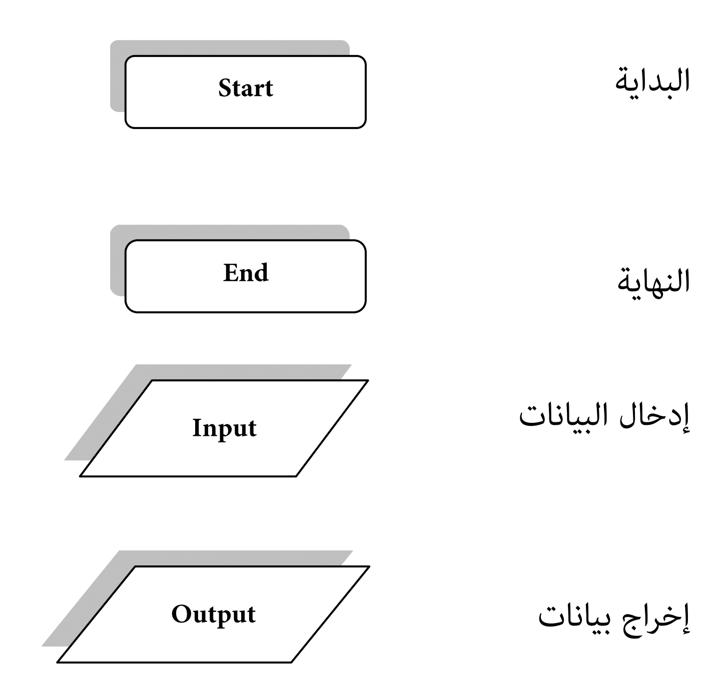
عند ترتيب خطوات العمل لتنفيذمهمة أو معالجة معنيةفان هذ المهمة تكون مفهوم وسهلةالتنفيذواعتمادا على ذلك يمكن تعريف الخوارزمية بأنهامجموعةمن الخطوات المنطقيةاللازمةلتنفيذمهمةأوعمل معين وتحتوي الخوارزميةعلى مدخلات تطبق

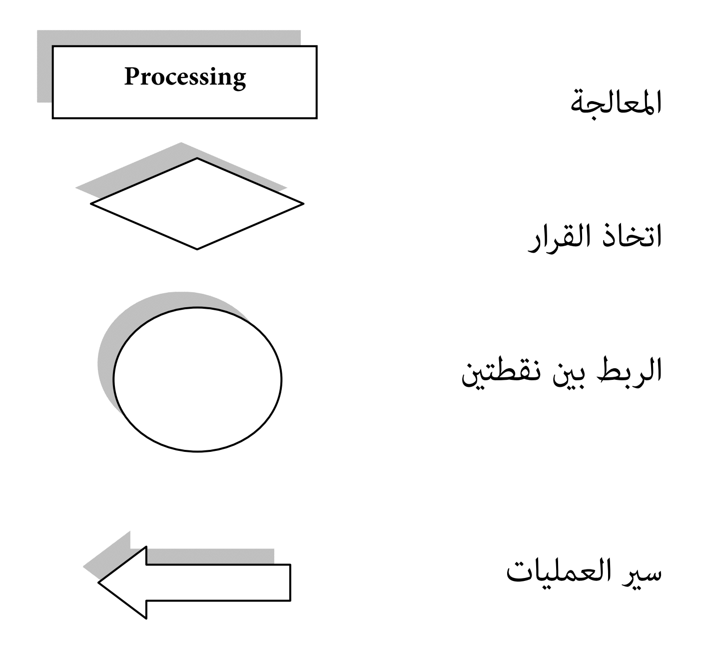
عليها عمليات معينة لكي نحصل على المخرجات ويمكن إعداد الخوارزمية باستخدام إحدى الوسيلتين :

أولا : الخطوات التسلسلية و تشمل :

١- البداية
٢- إدخال البيانات
٣- المعالجة
٤- إخراج البيانات
٥- النهاية

ثانيا : المخطط الانسيابي ويشمل طريقة استخدام رموز والمخططات والأسهم لتمثل العمليات المطلوبة حيث يبدأ المخطط الانسيابي ببداية البرنامج وينتهي بنهايته وما بينهما يمثل متن العمليات والإجراءات ومن هذه الرموز والأشكال نستعرض مايلي:



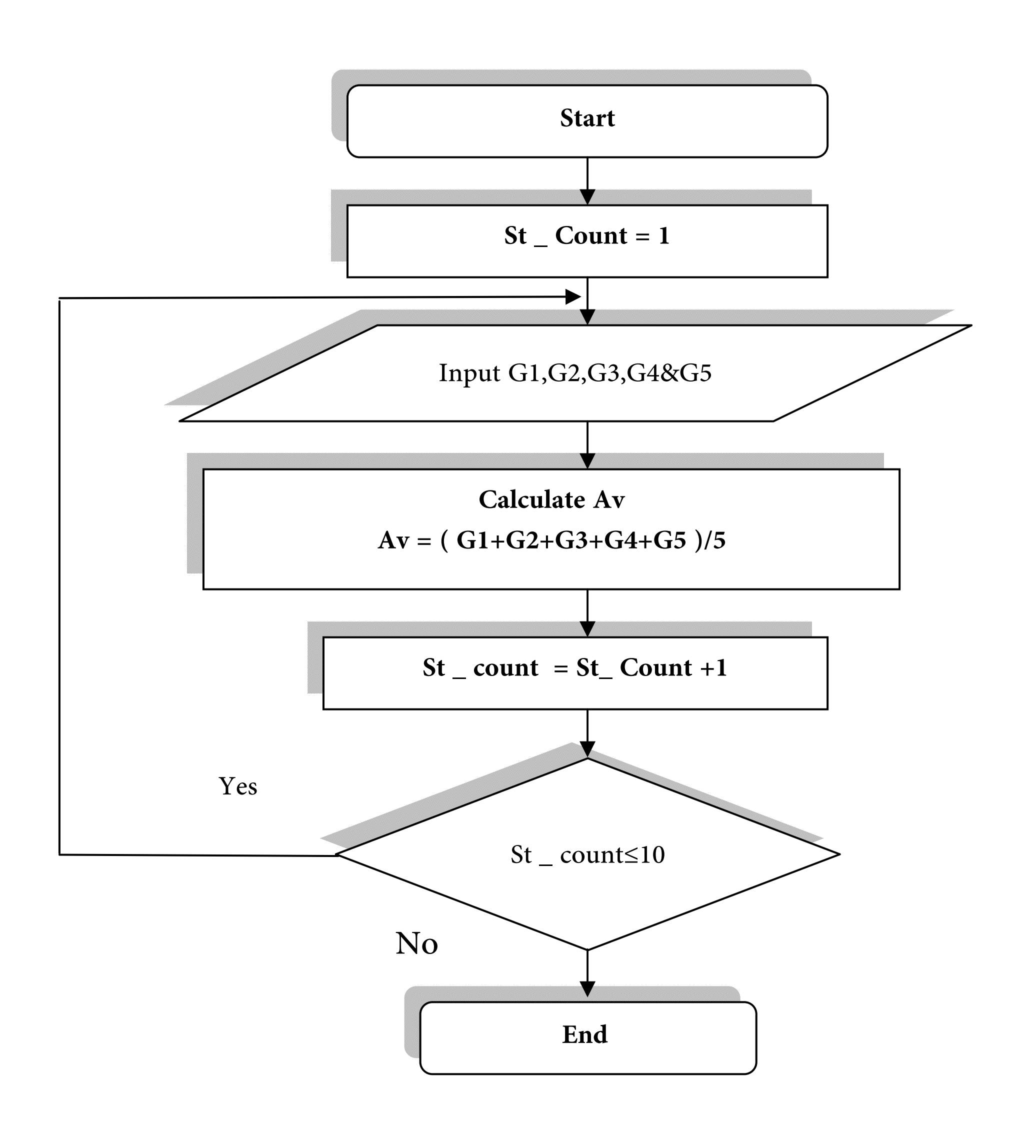


فلو أردنا استخدام الطريقتين السابقتين لحساب معدل علامات الطلبة على افتراض وجود خمسة علامات لكل طالب علما بان عدد الطلبة خمسة .

أولا: حل السؤال باستخدام الخوارزمية ذات الخطوات التسلسلية وكما يلي:

- تخصيص عداد للطلبة و إعطائه قيمة ١
- إدخال قيم العلامات G1,G2,G3,G4&G5 لكل طالب
- إيجاد المعدل $Av = (G1+G2+G3+G4+G5)/5$
- طباعة المعدل
- زيادة قيمة العداد بالعدد واحد و إعادة الحسابات
- النهاية عندما يصل العداد إلى عشرة

ثانيا: المخطط الانسيابي وكما يلي



٢.٧ حزمة MATLAB   MATLAB Package

أستخدمت حزمة MATLAB ظهرت وطبقت منذ عشرات السنين وبسبب أمكانيانها العالية تطورت وأنتشرت ودخلت في تطبيقات واسعة وفي الفصول القادمة سوف نستعرض أستخدامات هذه الحزمة وكيفية التعامل معها. والان سوف نطبق خطوات البرنامج في المثال السابق بأستخدام حزمة MATLAB فأنه بعد تشغيل الحزمة بأمكاننا وبشكل مبسط تنفيذ البرنامج بشكل مباشر ومن الشاشة الرئيسية حيث أولا ندخل العلامات لكل طالب على شكل صف، فلو كان الطالب الاول لديه العلامات (٨، ٨، ٩، ٩، ٧) فيتم أدخالها بأستخدام المتغيير G1 وكما يلي:

```
>> G1=[8 8 9 9 7]

        G1 =

        8    8    9    9    7
```

بأمكاننا حل ذلك بشكل مبسط وخطوة خطوة وذلك بجمع العلامات بأستخدام علامة الجمع الاعتيادية ويخزن الناتج في المتغير GG1 وكما يلي:

```
>> GG1=8+8+9+9+7

        GG1 =

        41
```

ثم بعد ذلك نقسم ناتج المجموع على عدد العلامات أي هنا نقسم على ٥ ونخزن الناتج في المتغيير GGG1 وكما يلي:

```
>> GGG1=GG1/5

        GGG1 =

        8.2000
```

وكذلك بالأمكان حل المثال بأستخدام الدوال الجاهزة لحزمة MATLAB ومنها دالة الجمع sum ففي البداية ندخل علامات الطالب وكما يلي:

```
>> G1=[8 8 9 9 7]

    G1 =

    8    8    9    9    7
```

بعد ذلك نستخدم الدالة sum لايجاد المجموع وكما يلي:

```
>> GG1=sum(G1)

    GG1 =

    41
```

ثم بعد ذلك نقسم ناتج المجموع على عدد العلامات أي هنا نقسم على ٥ وكما يلي:

```
>> GGG1=GG1/5

    GGG1 =

    8.2000
```

وكذلك بالأمكان حل المثال بأستخدام دالة أخرى وهي mean والتي تحسب المعدل مباشرة ففي البداية ندخل علامات الطالب وكما يلي:

```
>> G1=[8 8 9 9 7]

    G1 =

    8    8    9    9    7
```

بعد ذلك نستخدم الدالة mean لايجاد المعدل وكما يلي:

```
>> AVE=mean(G1)

    AVE =

    8.2000
```

ولاكمال حل المثال لخمسةطلبة نكررالعملية لخمسةمرات وحسب عددالطلبةوكما يلي:

ندخل علامات الطالب الاول ونحسب معدله

```
>> G1=[8 8 9 9 7]
G1 =
    8    8    9    9    7
>> AVE1=mean(G1)
AVE1 =
   8.2000
```

ندخل علامات الطالب الثاني ونحسب معدله

```
>> G2=[7 7 6 6 7]
G2 =
    7    7    6    6    7
>> AVE2=mean(G2)
AVE2 =
   6.6000
```

ندخل علامات الطالب الثالث ونحسب معدله

```
>> G3=[8 8 7 7 7]
G3 =
    8    8    7    7    7
>> AVE3=mean(G3)
AVE3 =
   7.4000
```

ندخل علامات الطالب الرابع ونحسب معدله

```
>> G4=[7 9 9 9 7]
G4 =
    7    9    9    9    7
>> AVE4=mean(G4)
AVE4 =
   8.2000
```

ندخل علامات الطالب الخامس ونحسب معدله

```
>> G5=[9 9 9 9 9]
G5 =
    9    9    9    9    9
>> AVE5=mean(G5)
AVE5 =
    9
```

بالإمكان حل المثال بوضع كافة علامات للطلبة على شكل مصفوفة حيث تكون علامات كل طالب موجودة في عمود من المصفوفة فعلامات الطالب الأول في العمود الأول وعلامات الطالب الثاني في العمود الثاني وهكذا. ثم أيجاد المعدل لكل عمود وهو يمثل معدل علامات الطالب وهكذا وكما يلي:

```
>> G=[8 7 8 7 9;8 7 8 9 9;9 6 7 9 9;9 6 7 9 9;7 7 7 7 9]
G =        8    7    8    7    9
           8    7    8    9    9
           9    6    7    9    9
           9    6    7    9    9
           7    7    7    7    9
```

```
>> AVE=mean(G)

AVE =

   8.2000    6.6000    7.4000    8.2000    9.0000
```

وسوف نتطرق لاحقا في الفصول القادمة بشكل أكثر تفصيلا عن أستخدام حزمة MATLAB في البرمجة وحل المسائل الرياضية والاحصائية وكذلك في أستخدام وتطبيق برامج أكثر تعقيدا.

الفصل الثالث

المفاهيم الأساسية لحزمة MATLAB

Basic Concepts of MATLAB Package

٣.١ مقدمــة	Introduction
٣.٢ مراحل التحميل الأساسي	Basic Installation Procedure
٣.٣ متطلبات النظام	System Requirements
٣.٤ مواصفات النظام	System Specifications
٣.٥ محتويات النظام	System Contents
٣.٦ تقنيات البرمجة	Programming Techniques
٣.٧ طرق البرمجة	Programming Methods
٣.٨ العمليات الرياضية	Mathematical Operations
3.9 المصفوفات	Matrices
٣.١٠ البرامج الإحصائيةَ في الحزمة	Statistical Toolbox
٣.١١ الاستخدامات الإحصائية	Statistical Applications

٣.١ مقدمــة Introduction

حزمة MATLAB عبارة عن مجموعة من البرامج والوظائف لاداء مهام وعمليات مختلفة وقد كان الهدف الرئيسي للأصدارات الاولى هو لتسهيل تنفيد أعمال ومهام بأستخدام التطبيقات الرياضية. ظهرت حزمة MATLAB منذ فترة طويلة وبالتحديد في الثمانينيات حيث تم تطبيقها على بيئة أنظمة التشغيل MS-DOS حيث كانت تقتصر على أوامر وتطبيقات رياضية وأحصائية بسيطة ونظرا لكفائة الحزمة ودخولها في مجالات واسعة بدأت تشق طريقها بقوة بين البرمجيات الجاهزة. تطورت حزمة MATLAB مع تطور البيئة التشغيلية للحاسوب إلى أن ظهرت بيئة النوافذ WINDOWS حيث تم تطوير الحزمة لكي تتلائم مع هذه البيئة، ومن تلك الفترة ولحد الآن ظهرت نسخ مختلفة من حزمة MATLAB وكل نسخة جديدة تظهر تضاف إليها تقنيات وأدوات جديدة وأن آخر نسخة ظهرت والتي سنعتمد عليها في هذا الكتاب هي النسخة السابعة Version 7 حيث أضافت تطبيقات واسعة في مجالات مختلفة. أضافة الى ما تقدم فأن حزمة MATLAB تمتاز بمواصفات كثيرة من أهمها ما يلي:

١. تحتوي حزمة MATLAB على دوال الجاهزة بالإمكان استخدامها في الحزمة لتنفيذ المهام المطلوبة.

٢. تحتوي حزمة MATLAB على لغة برمجة الحزمة بالإمكان استخدامها كلغة برمجة لتنفيذ المهام المطلوبة.

٣. تحتوي حزمة MATLAB على حقل المحاكاة بالإمكان استخدامه لتصميم وتنفيذ المهام المطلوبة.

٤. تحتوى الحزمة على كمية كبيرة من التطبيقات المتفرعة التي يمكن الاستفادة منها في مجالات مختلفة.

٥. لهذه الحزمة القابلية والقدرة العالية على التعامل وحل كافة المشاكل والمهام المطلوبة بأستخدام أمكانياتها العالية.

٦. لهذه الحزمة المرونة العالية في التكيف والتوافق والتعامل مع لغات البرمجة والبيئات الأخرى لتسهيل العمل.

٧. لهذه الحزمة المرونة في تحديث النسخة عن طريق الانترنت وإضافة كافة الأمور والاصدارات الجديدة.

٨. لهذه الحزمة القابلية والقدرة العالية على التعامل مع الرسومات والمخططات والتكييف حسب نوعية العينة وحجمها.

٣.٢ مراحل التحميل الأساسي Basic Installation Procedure

عملية تحميل ونقل حزمة MATLAB إلى الحاسوب عملية بسيطة وغير معقدة وبأمكانك القيام بهذه المهمة بشكل سليم، ولاجل ذلك يجب تهيئة كافة المتطلبات المادية والبرمجية للحاسوب لكي يكون قادرا على تحميل هذه الحزمة. وسوف نستعرض الخطوات الأساسية لتحميل وتركيب حزمة MATLAB على جهاز الحاسوب.

١. الخطوة الأولى: قبل البدء بالتنصيب

قبل البدء بتنصيب حزمة MATLAB يجب أن تحصل على كلمة السر الشخصية personal Lice use password (PLP) والمكونة من خمسة مقاطع وكل مقطع من خمسة أرقام أو رموز وهو الذي يحدد ويعرف المنتج الذي يسترخصه، وعند شرائك للمنتج بإمكانك الحصول على كلمة السر عن طريق البريد الالكتروني.

٢. الخطوة الثانية: البدء بالتنصيب

ادخل القرص ١ واضغط على تنصيب Install، تم بعدها next حيث تبدأ الحزمة بالتهيؤ للتحميل.

٣. الخطوة الثالثة: إدخال التعريف

يبدأ البرنامج بطلب معلومات عن المستخدم وهي ادخل اسمك واسم الشركة وكلمة السر في المعلومات المرخصة وبعدها اضغط على حيث يبدأ البرنامج بالتنصيب.

٤. الخطوة الرابعة: الموافقة على شروط البرنامج

ولاستمرار تنصيب البرنامج تظهر شروط البرنامج ولإكمال ذلك اضغط على نعم yes ثم اضغط على next.

٥. الخطوة الخامسة: أختيار نوع التنصيب

البرنامج يخير المستخدم بأختيار أحد النوعين للتنصيب أما مثالي أو حسب الطلب ثم اضغط على next.

٦. الخطوة السادسة: تجديد موقع خزن البرنامج

البرنامج يحدد الموقع الذي تخزن به الحزمة وكذلك من حق المستخدم أن يختار الموقع الذي يخزن به الحزمة في الحاسوب ثم نضغط على next.

٧. الخطوة السابعة: تحديد الخيارات المطلوب نقلها

تحدد الأمور والبرامج التي يراد نقلها وهذا فقط يعمل بالنسبة إلى الخيار حسب الطلب ثم نضغط على next.

٨. الخطوة الثامنة: تأكيد الخيارات

قبل البدء بنقل الملفات تظهر على الشاشة ملخص عن كل التنصيب ولاستمرار التنصيب اضغط على تنصيب Install. حيث يبدأ بالنقل وعند إكمال القرص يطلب القرص الثاني وهكذا.

٩. الخطوة التاسعة: اقرأ ملاحظات ترتيب المنتج

هذه معلومات أو أرشادات للمستخدم وتتضمن معلومات عن المنتج ومعلومات عن التحديث وبعدها نضغط على next.

١٠. الخطوة العاشرة: إكمال التنصيب

بعد إكمال التنصيب تظهر على الشاشة هل تريد تشغيل البرنامج عندها أشر على START MATLAB واضغط على نهاية FINISH.

١١. الخطوة الحادية عشر: بعد التنصيب

عند أكمال عملية التنصيب وللتأكد من أن العمل قد أكمل بالشكل الصحيح حيث تظهرعلامة MATLAB على جهازك وبالامكان ملاحظة ذلك. ولتشغيل الحزمة نضغط على هذه العلامة حيث تظهر شاشة الحزمة وكما موضحة في الشكل ٤.١ .ولتشغيل وطلب المساعدة في أي موضوع ضمن الحزمة نضغط على Help حيث تظهر لنا الشاشة كما موضحة في الشكل ٣.٢.

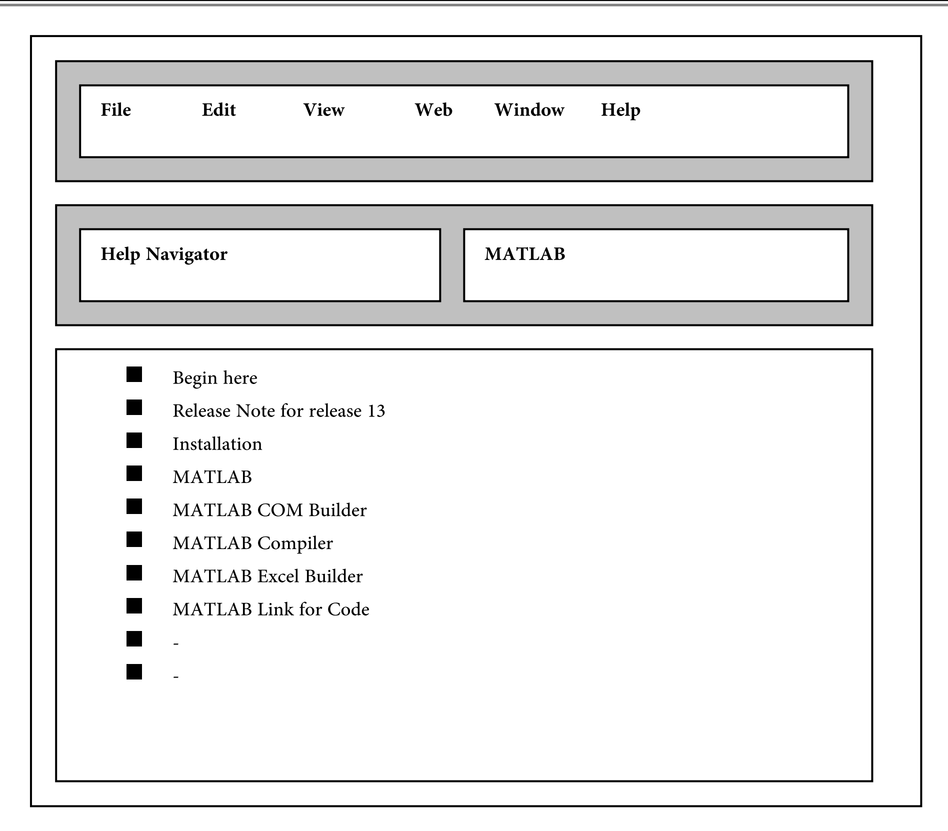
File Edit View Web Window Help

Using Toolbox Path Cache. Type "help toolbox_path_cache" for more info.

To get started, select "MATLAB Help" from the Help menu.

>>

شكل ٣.١ الشاشة الرئيسية لحزمة MATLAB



شكل ٣.٢ الشاشة الثانوية الخاصة بالمساعدة في حزمة MATLAB

٣.٣ متطلبات النظام System Requirements

لأجل تنصيب حزمة MATLAB والعمل بها بشكل جيد ومقنع وتنفيذ كافة الدوال والتطبيقات المتوفرة بها هناك مجموعة من المتطلبات يجب تحقيقها وتوفيرها وهي:

١. حاسوب بنتيوم ٣ صعوداً يتضمن سعة معقولة لنقل الملفات ويحتوي على ذاكرة ٥١٢ ميغابايت ويحتوي على قارئ الأقراص المدمجة.

٢. ينصب نظام تشغيل حديث ويفصل WINDOWS XP.

٣. ينصب مستكشف الانترنت Microsoft Internet Explorer 4.0.

٤. ينصب قارئ الادوبي Adobe Acrobat Reader 3.0

5. ينصب بروتوكول الانترنيت TCP/IP
6. ضرورة وجود طرف USB.
7. تكون مواصفات الرسومات بدقة 32 bit.
8. نظام التشغيل يكون مدعم بكارت الرسومات وكارت الصوت.

٣.٤ مواصفات النظام System Specifications

تقوم حزمة MATLAB بأداء مهام وأعمال واسعة بمختلف الاختصاصات ولها مرونة عالية في التعامل مع المشاكل وحلها سواء بطريقة البرمجة أو بطريقة الحزم الجاهزة. إضافة إلى ذلك أنها تضيف خصوصية وميزة مهمة بتطورها مع تطور التقنيات والتكنولوجيا وبيئات التشغيل والتكييف المستمر مع هذه التقنيات والبيئات. حزمة MATLAB لها القابلية والقدرة على التعامل والتوافق مع لغات البرمجة الأخرى من حيث الأداء واستخدام الدوال والأوامر والربط والتكييف مع هذه اللغات. حزمة MATLAB تتضمن لغة عالية الأداء تستخدم للحسابات التقنية والفنية حيث تقوم بتكامل الحسابات وإظهارها ودمجها في بيئة سهلة الاستعمال في التعامل مع المشاكل وإيجاد الحلول المناسبة وتمثيلها بدلالات رياضية واضحة ومفهومة وسهلة التطبيق.

تتضمن حزمة MATLAB الاستعمالات المثالية التالية ما يلي:

1. الرياضيات والحسابات.
2. تطوير الخوارزميات.
3. استلام البيانات.
4. النمذجة والمحاكاة.
5. تحليل البيانات وعرضها.
6. التطبيقات العالية والهندسية.
7. واجهات المستخدم.

٣.٥ محتويات النظام System Contents

تحتوي حزمة MATLAB على تطبيقات واسعة وبأختصاصات مختلفة وأمثلة عملية في كافة حقول المعرفة وبالأمكان أضافة أو تطوير أي دالة أو مهمة لتلائم ذلك العمل وبطرق مختلفة حيث يصبح هذا التطوير جزء من بيئة العمل وأن حزمة MATLAB عبارة عن حزمة ضخمة تحتوي على أدوات ودوال كثيرة لتأدية مهام وأعمال واسعة وعلى مستوى تطبيقات مختلفة ومن الأجزاء الرئيسية لهذه الحزمة ما يلي:

١. بيئة التطوير Development Environment

وهي مجموعة الأدوات التي تساعد في استخدام الدوال والملفات الخاصة بحزمة MATLAB والتي تؤدي بتكاملها الى قوة وكفاءة الحزمة.

٢. مكتبة الدوال الرياضية Mathematical Function Library

وهي كم هائل من الخوارزميات الحسابية والرياضية المتعلقة بالوظائف الأولية والخاصة ويمكن أستخدام المساعدة في الدلالة عليها وأستخدامها.

٣. لغة الحزمة MATLAB Language

وهي لغة المستوى العالي مع كافة الدوال والسيطرة على سير العمليات والبيانات وخواص البرمجة وتطبيقاتها وتمتاز بالسهولة والمرونة وسرعة التنفيذ.

٤. الرسومات Graphics

للحزمة وسائل شاملة لتبويب وعرض البيانات على شكل رسومات ومخططات بيانية وبأنواعها المختلفة حيث بأمكانك رسم أي متغيير أو دالة أذا تم تعريف عناصرها.

٥. تطبيق واجهة البرنامج Application Program Interface

وهي عبارة عن مكتبة تمكنك من كتابة البرنامج باللغات الأخرى كلغة السي ويمكن التعامل معها بحزمة MATLAB .

٦. المساعدة Help

للحزمة مساعدة فعالة جداً من ناحية الدوال وتطبيقاتها ومن ناحية اللغة واستخدامها حيث بأمكانك طلب المساعدة عن تطبيق أو أستخدام أي دالة ضمن بيئة العمل.

٣.٦ تقنيات البرمجة Programming Techniques

يمكن التعامل مع لغة البرمجة بأستخدام حزمة MATLAB بكل سهولة وأنها تستخدم أوامر ودوال بسيطة وذات مرونة عالية، أضافة الى ذلك فأن المساعدة تدلك على الطريق الصحيح لاستخدام هذه اللغة وكيفية تركيب الدوال فيها. أما من حيث البرمجة فهي تشبه بشكل كبير لغة السي المبسطة وبفارق بسيط هو أنك لاتحتاج الى تعريف المتغيرات بأستخدام هذه اللغة. ويمكن أستخدام الواجهة الرئيسية لتطبيق وتنفيذ أوامر قليلة وأظهار النتائج بشكل مباشر أما أذا أردت كتابة برامج معقدة فبأمكانك أستخدام بيئة الملفات M للخزن والتنفيذ بعد ذلك.

لغة البرمجة باستخدام حزمة MATLAB هي لغة عالية المستوى تتضمن تراكيب بيانات أساسها المصفوفة وأنواع بياناتها الداخلية الخاصة ودليل شامل من الوظائف وبيئة لتطوير أي وظيفة أو مهمة ولها القدرة على استلام وإرسال البيانات والمعلومات إلى أنواع مختلفة من ملفات البيانات وكذلك إمكانيات البرمجة الموجهة للكائنات وعمل واجهات للتقنيات الخارجية والتوافق مع لغات البرمجة الأخرى. وفيما يلي بعض من الميزات والتقنيات البرمجة:

١. هياكل البيانات data Structures.

٢. أنواع البيانات Data types.

٣. مكونات البرنامج الأساسية Basic program components

٤. برمجة الملف M-file programming.

٥. أنواع الوظائف Types of functions

٦. إرسال واستلام البيانات Data Import and Export.

٧. معالجة الخطأ Error Handling.

٨. الإصناف والأجسام Classes and Objects.

٩. جدولة تنفيذ البرنامج بالمؤقتات Scheduling Program Execution with Timers.

١٠. تحسين أداء واستخدام الذاكرة Improving performance and Memory Usage.

١١. برمجة النصائح Programming Tips.

3.7 طرق البرمجة Programming Methods

هناك طريقتين للبرمجة باستخدام حزمة MATLAB واختيار اي منهما يعتمد على المستخدم وعلى حقل التطبيق. ويجب أن نوضح هنا أنه بعد تشغيلك خزمة MATLAB تظهر شاشة الحزمة وفي أعلى اليسار تظهر العلامة (>>) التي من خلالها يمكن ادخال المعلومات.

- **البرمجة المباشرة**

وعادة يستخدمها المبتدئين عندما يكون عمله عبارة عن خطوات قليلة ويريد التأكد من برنامج صغير الحجم وتعتمد على ادخال كل المتغيرات المطلوبة من بيانات ومتغيرات ودول بشكل مباشر على الواجهة الرئيسية حيث يتم تنفيذ البرنامج على شكل خطوات وتنفذ كل خطوة عند تنفيذ الامر بادحالها وهنا تبقى كل قيم المتغيرات مخزونة اذا لم يتم تغييرها بقيم اخرى فلو فرضنا انه اردنا اعطاء المتغير X القيمة ٥ وبدون أظهار القيمة على الشاشة فنكتب الامر كما يلي:

```
>>  X = 5; ↵
>>
```

ولو اردنا اظهار النتيجة على الشاشة لنفس المتغير نكتبه كما يلي :

```
>> X = 5 ↵
   X =
     5
```

وعند ذلك فان المتغير X قد اخذ القيمة خمسة وتبقى مخزونة معه مالم تتغير ولو لردنا استخدامها مثلا لايجاد قيمة Y التي تساوي مرع قيمة X فاننا نكتبها كما يلي:

```
>> Y  = X * X ↵
   Y =
     25
```

- **البرمجة الغير مباشرة**

وعادة يستخدمها المتخصصين و المبرمجين عند تصميم برنامج بخطوات كثيرة حيث يكون حجم البرنامج كبير، وتعتمد بالاساس على فتح ملف يعطى اسم معين مثلا PROJ.M وتكتب فيه كل خطوات البرنامج المطلوبة وبعدها تخزينه في الموقع WORK يتم تنفذ البرنامج باستحدام الامر RUN حيث يعطيك النتيجة المطلوبة وبما انه مخزون لديك في الموقع WORK فبامكانك تعديله وتغييره في اي وقت وحسب المتطلبات وفي النهاية سوف تصبح لديك داله جاهزة اسمها PROJ يمكن استدعائها عن طريق الواجهة الرئيسية وكما يلي:

FILE → OPEN → WORK → PROJ

وكذلك يمكن استدعائها لكتابة مايلي على الواجهة الرئيسية

```
>> EDIT PROJ
```

ويمكن تنفيذ البرنامج من الواجهة الرئيسية بكتابة PROJ ثم الضغط على ألأدخال

```
>> PROJ
```

٣.٨ العمليات الرياضية Mathematical Operations

بعد ان استعرضنا كيفية التعامل عند حزمة MATLAB سنستعرض اهم العمليات الرياضية التي بامكانك استخدامها وهي :

- الجمع ADDITION ويرمز لها (+) وكما يلي:

```
>> X = 7;
>> Y = 5;
>> Z = X + Y
   Z = 12
```

- الطرح SUBTRACTION ويرمز لها بالرمز (-) وكما يلي:

```
>> X = 7;
>> Y = 5;
>> Z = X - Y
   Z = 2
```

- الضرب MULTIPLICATION ويرمز لها بالرمز (*) وكما يلي:

```
>> X = 7;
>> Y = 5;
>> Z = X * Y
   Z = ٣٥
```

- ضرب الصفوف ARREY MULTIPLICATION ويرمز لها بالرمز (.*) ويطبق على كل عنصر لوحده مع العنصر المقابل وكما يلي:

```
>> X = [2    5];
>> Y = [3    8];
>> Z = X. * Y
   Z = 6     40
>> X1 = [1    2,3    4];
>> Y1 = [2    3,4    5];
>> Z1 = X1 . * Y1 ↵
   Z1= 2     6
       12    20
```

- القسمة RIGHT DIVISION ويرمز لها بالرمز (/) وكما يلي:

```
>> X = 7;
>> Y = 5;
>> Z = X / Y
   Z = 1.4
```

- قسمة الصفوف ARREY RIGHT DIVISION ويرمز لها بالرمز (./) ويطبق على كل عنصر لوحده مع العنصر المقابل وكمايلي:

```
>> X = [2    5];
>> Y = [3    8];
>> Z = X . / Y
   Z = 0.6667     0.6250
>> X1 = [1    2,3    4];
>> Y1 = [2    3,4    5];
```

```
>> Z1 = X1 . / Y1 ↵
   Z1= 0.5000        0.6667
         0.7500        0.8000
```

- القسمة اليسرى LEFT DIVISION ويرمز لها بالرمز (\) وكما يلي:

```
>> X = 7;
>> Y = 5;
>> Z = X \ Y
   Z = 0.7143
```

- قسمة الصفوف اليسرى ARREY LEFT DIVISION ويرمز لها بالرمز (.\) ويطبق على كل عنصر لوحده مع العنصر المقابل وكما يلي:

```
>> X = [2      5];
>> Y = [3      8];
>> Z = X .\ Y
   Z = ١.٥        ١.٦

>> X1 = [1      2,3     4];
>> Y1 = [2      3,4     5];
>> Z1 = X1 .\ Y1 ↵
   Z1= ٢.٠٠0        ١.٥٠٠٠
         ١.٣٣٣٣        1.2500
```

- الاسس او الرفع الى القوة RAISED TO A POWER ويرمز لها بالرمز (^) وكما يلي:

```
>> X = 3;
>> Y = 4;
>> Z = X ^ Y
   Z = 81
```

وكذلك يمكن كتابة الطريقة الأسية بشكل أخر وكما يلي:

```
>> X = 3;
>> Y = 4;
>> Z = POWER (X,Y) ↵
   Z = 81
```

- رفع الصفوف الى القوة ARREY RAISED TO A POWER ويرمز لها بالرمز (^.) ويطبق على كل عنصر لوحده مع العنصر المقابل وكما يلي:

```
>> X = [2     3];
>> Y = [5     4];
>> Z = X .^ Y↵
   Z = 32       81
>> X1 = [1     2,3     4];
>> Y1 = [2     3,4     5];
>> Z1 = X1.^Y1 ↵
   Z1= 1             8
         81          1024
```

- COMPLEX CONJUGATE TRANSPOSE ويرمزله بالرمز(’) وكما يلي:

```
>> X = 2+3i
>> Y = X’
 Y = 2 – 3i
```

- REAL TRANSPOSE ويرمز له بالرمز (’.) وكما يلي:

```
>> X = 2+3i,
>> Y = X.’
 Y = 2 + 3i
```

3.9 المصفوفات Matrices

تعتبر المصفوفات من المواضيع المهمة التي ترسل في الامور الرياضية والاحصائية وان حزمة MATLAB بامكانها اجراء كافة العمليات الحسابية التي تطبق على المصفوفات ومنها:

- تكوين المصفوفة

يمكن انشاء مصفوفة على شكل عمود واحد وكما يلي:

```
>> X = [3; 5; 8]
>> X1 =
        3
        5
        8
```

ويمكن انشاء مصفوفة على شكل سطر واحد وكما يلي:

```
>> X = [ 1   2   7]
>> X2 =
       1  2  7
```

وكذلك يمكن انشاء مصفوفة بعدة صفوف وعدة اعمدة وحسب مامطلوب وكما يلي:

```
>> X = [ 1  2  3; 4  5  6; 7  8  9]
>> X =
        1    2    3
        4    5    6
        7    8    9
```

وبالإمكان إنشاء مصفوفة وتحديد عدد الاعمدة والصفوف وان قيم العناصر تكون بشكل عشوائي باستخدام الدالة rand وفي المثال التالي هذه المصفوفة مكونة من ثلاثة صفوف واربعة اعمدة:

```
>> X = rand [ 3,4]
   X =
        0.3    0.5    0.8    0.7
        0.2    0.7    0.5    0.5
        0.3    0.3    0.3    0.4
```

بالامكان أستخدام نفس الدالة السابقة لايجاد قيم بأعداد صحيحة وكما يلي:

```
>> X = fix (10 * rand (3,4))
   X =
        3    5    8    7
        2    7    5    5
        3    3    3    4
```

٣.١٠ البرامج الإحصائيةَ في الحزمة Statistical Toolbox

البرامج الإحصائية الموجود مَع حزمة Matlab ،هو مجموعة الأدوات الإحصائيةِ بنيت على بيئةِ استعمال الحسابات العددية باستخدام Matlab . صندوق العُدّة يساند تشكيلة واسعة من المهامِ الإحصائيةِ المشتركةِ، مِنْ توليد العددِ العشوائي، إلى ثبوت البيانات وعرضها، إلى تصميمِ التجاربِ والسيطرة على المتغيرات العشوائية. تزود البامج الاحصائية في الحزمة بأثنان مِنْ أصنافِ الأدواتِ هي:

أولا: احتمالية بناء الوظائفِ الإحصائيةَ

ثانيا: أدوات تفاعلية تخطيطية

إنّ الصنفَ الأولَ للأدواتِ يمثل الوظائفِ التي بالامكان ستدعائها من خلال طريقة كتابة الأوامر أَو مِنْ التطبيقات الخاصة والعديد مِنْ هذه الوظائفِ موجودة في ملفات من نوع m، وهي عبارة عن بيانات Matlab التي تُطبّق خوارزميات إحصائية متخصصةً.يمكنك أَنْ تستخدم الرموز الخاصة بحزمة Matlab لهذه الوظائفِ باستعمال أسم الوظيفة function_name ويمكنك أَنْ تُغيّرَ الطريقَ أيّ الوظيفة بنَسْخ وتبديل أسم الملف، ثمّ تعدّل نسختَكَ. وكذلك يمكنك أَنْ توسع البرامجيات أيضاً بإضافة ملفاتك الخاصة التي تدل على وظائف جديدة.

أما الصنف الثانيَ، يُزوّدُ الحزمة بعدد مِنْ الأدواتِ التفاعليةِ التي تَتيح لك أن تَدْخلُ العديد مِنْ الوظائفِ خلال واجهة المستعمل بالرسوم (graphical user interface).

والمثال التالي يوضح كيفية أستخدام المساعدة في معرفة تطبيق الدالة abs وهي تعطي القيمة المطلقة للمتغيرات وكما يلى:

```
>> help abs

   ABS    Absolute value.

   ABS(X) is the absolute value of the elements of X. When
X is complex, ABS(X) is the complex modulus (magnitude) of the
elements of X.

    See also SIGN, ANGLE, UNWRAP.
```

```
Overloaded methods
   help sym/abs.m
   help iddata/abs.m
>>
```

٣.١١ الاستخدامات الإحصائية Statistical Applications

صندوق الحزمة الإحصائيةَ لَهُ أكثر مِنْ ٢٠٠ ملف، وأهم المواضيع المساند له تتضمن الحقول التاليةِ:

١. توزيعات إلأحتماليةِ Probability Distributions

َيدْعمُ صندوقُ الحزمة الإحصائيةَ ٢٠ توزيعَا إحتماليا ولكُلّ توزيع هناك خمس وظائفِ مرتبطةِ به وهذه الوظائف هي:

- وظيفة كثافةِ إلأحتماليةِ (pdf)
- وظيفة التوزيعِ المتراكمةِ (cdf)
- وظيفة المعكوسة لوظيفةِ التوزيعِ المتراكمةِ
- وظيفة مولّدِ البياناتِ العشوائي للمتوسط والتباين
- وظيفة حسابات التخميناتِ وفتراتِ الثقةِ.

٢. الإحصاء الوصفي Descriptive Statistics

يُزوّدُ البرامج الإحصائية لحزمة MATLAB وظائفَ وَصْف البيانات وتوزيعها وتبويبها وعرضها على شكل رسوم بيانية مع تقدير القيم المفقودة من سلسلة البيانات.

٣. النماذج الخطية Linear Models

في حقل النماذج الخطية، يمكن الاستفادة من البرامج الاحصائية في الحزمة في تحليل التباين بأتجاه واحد وأتجاهين أو أكثر (ANOVA) أضافة الى تحليل التغاير (ANCOVA) وتحليل الانحدار الخطي والمتعدد وعمليات التنبؤات المستقبلية.

٤. النماذج اللاخطّية Nonlinear Models

يمكن الاستفادة أيضا من البرامج الاحصائية في الحزمة في تقدير المعالم للنماذج الغير خطية وتحديد فترات الثقة ومعرفة التوزيعات اللامعلمية مع أيجاد التوقعات لمعالمها.

٥. إختبارات فرضيةِ Hypothesis Tests

كذلك يُزوّدُ صندوقُ العُدّة الإحصائية الوظائفَ التي تعمَلُ الإختباراتُ الأكثر شيوعاً مِنْ الفرضيةِ -- إختبارات T، إختبارات Z ، وإختبارات اللامعلمية، وإختبارات التوزيعِ الاخرى.

٦. الأحصائيات متعددة المتغيرات Multivariate Statistics

يَدْعمُ صندوقُ العُدّة الإحصائية الطرقَ الإحصائية في متعددةِ المتغيرات، بضمن ذلك تحليلِ المكوّنات الرئيسية لنماذج متعددة المتغيرات ويضاف الى ذلك التحليل العاملي، أضافة الى تحليل التباين بأتجاه واحد أو أتجاهين أو أكثر أضافة الى أستخدام نماذج التحليل العنقودي .

٧. الرسومات الإحصائية Statistical Plots

تضيف البرامج الإحصائية برنامج صندوق الرسم، رسم التوزيع الطبيعي ، رسم إحتمالية Weibull ، لوحات مراقبة، الرسوم البيانية في حزمة Matlab. هناك دعم أيضاً لتركيب وتنبؤِ المنحنى المتعدّد الحدودِ. هناك وظائف لاعداد رسومات البيانات المنتشرة أو مصفوفاتِ الرسومات المنتشرة للبياناتِ المُجَمَّعةِ، لمعرفة توزيع البيانات والتفاعلات بين المتغيرات في نماذج الانحدار الخطي.

٨. سيطرة العمليةِ الإحصائيةِ Statistical Process Control (SPC)

لسيطرةِ العمليةِ الإحصائيةِ، تزود البرامج الإحصائيةَ بوظائفَ تَخطيط الوحاتِ للمراقبة والسيطرة النوعية.

٩. تصميم التجاربِ Design of Experiments (DOE)

تدعم البرامج الإحصائياتَ تصاميمَ متعددة لعرض أعداد التصاميم التجريبية مثل تصميم العاملي العاملي والتصاميم العشوائية التامة وتصميم المجاميع العشوائية.

١٠. نماذج سلسلة ماركوف Hidden Markov Models

تحتوي البرامج الإحصائية في الحزمة بوظائفَ لتَحليل نماذجِ Markov. والتي يمكن من خلالها توليد البيانات حيث يتم أيجاد القيم التخمينية لمعالم النموذج مع تقدير الحدود العليا والدنيا لمعرفة حدود تقدير المعالم المسموح بها مع معرفة أفضل قيمة تقديرية للمعالم المقدرة.

الفصل الرابع

تطبيقات حزمة MATLAB في الإحصاء الوصفي

Applications of Matlab in Descriptive Statistic

٤.١ مقدمة Introduction

سيتضمن هذا الفصل توضيح تطبيقات حزمة MATLAB في موضوع الإحصاء الوصفي للبيانات مع أسئلة تطبيقية لكل حالة. مقياس النزعة المركزيِة (Measures of Central Tendency) ويتم حساب المتوسطات ، مثل الوسطِ الحسابي والوسطِ الهندسي والوسيط والمنوال وغيرها مع أمثلة لها. أضافة الى مقاييس التشتت (Measures of Dispersion) ويتم قياس الانحراف المعياري والتباين والمدى مع أمثلة لها. تقدير القِيَمِ المفقودةِ (Missing Values NaNs) سيتم توضيح كيفية أستحدام حزمة MATLAB لتقدير القِيَمُ المفقودة في البيانات. كما سيتضمن الفصل تصنيف البيانات وفق مجاميع (Grouped Data) في هذا الجزء سيتم أستحدام حزمة MATLAB لتصنيف البيانات وفق مواصفات معينة بحيث تكون كل مجموعة أكثر تجانسا مما يسهل على الباحثين وصف وتحليل تلك البيانات. أضافة الى وصف وعرض البيانات بالأشكال (Graphical Descriptions) حيث سيتم أستخدام حزمة MATLAB في وصف وعرض البيانات بالاشكال التخطيطية المختلفة. ثم بعد ذلك تصنيف البيانات بطريقة The Boot strap وهي طريقة تصنيف البيانات وفقا لمواصفات تلك البيانات وسوف نستعرضها بأسخدام حزمة MATLAB.

٤.٢ مقاييس النزعة المركزيِة Measures of Central Tendency) (Location

مقياس النزعة المركزية تشمل الوسط الهندسي، الوسط التوافقي، الوسط الحسابي، الوسيط، المنوال، والمعدل الحسابي وكل هذه المقاييس لها تطبيقات واسعة في نظم المعلومات وفي أتخاذ القرارات حيث أن هذه المقاييس تعطي مؤشرات ومعلومات كثيرة حول مواصفات العينة. سنتناول عدد من مقاييس النزعة المركزية في هذا الجانب لغرض التعرف على كل الاوساط التي يمكن أستخدامها لغرض البحث:

١. الوسط الهندسي (Geometric mean) ويمكن أيجاده بأستخدام الدالة geomean

يحسب الوسط الهندسي بأستخدام حزمة MATLAB وذلك بتطبيق الدالة geomean ومثال على ذلك نعطي المتغير x عبارة عن صف من أربعة قيم ثم نجد الوسط الهندسي بأستخدام الدالة geomean وكما يلي:

```
>> x=[1 3 6 2]

    x =

    1    3    6    2

>> geomean(x)

    ans =

    2.2136
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٤*٤ من القيم ثم نطبق نفس الطريقة لايجاد الوسط الهندسي حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

    y =

    1    2    3    6

    2    6    3    6

    2    6    6    8

    6    6    6    8

>> geomean(y)

    ans =

   2.0000    3.7226    3.8337    6.2603
```

٢. الوسط التوافقي (Harmonic mean) ويمكن أيجاده بأستخدام الدالة <u>harmmean</u>

يحسب الوسط التوافقي بأستخدام حزمة MATLAB وذلك بتطبيق الدالة <u>harmmean</u> ومثال على ذلك نعطي المتغير x صف من أربعة قيم ثم نجد الوسط التوافقي بأستخدام الدالة harmmean وكما يلي:

```
>> x=[1 3 6 2]

     x =

     1    3    6    2

>> harmmean(x)

     ans =

     1.9200
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٤*٤ من القيم ثم نطبق نفس الطريقة لايجاد الوسط التوافقي حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

     y =

     1    2    3    6

     2    6    3    6

     2    6    6    8

     6    6    6    8

>> harmmean(y)

     ans =

     1.7778    3.6286    3.6923    6.0000
```

٣. الوسط الحسابي (Arithmetic mean) ويمكن أيجاده بأستخدام الدالة mean

يحسب الوسط الحسابي بأستخدام حزمة MATLAB وذلك بتطبيق الدالة mean ومثال على ذلك نعطي المتغير x صف من أربعة قيم ثم نجد الوسط التوافقي بأستخدام الدالة mean وكما يلي:

```
>> x=[1 3 6 2]

        x =

        1    3    6    2

>> mean(x)

        ans =

        2.5000
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٤*٤ من القيم ثم نطبق نفس الطريقة لايجاد الوسط الحسابي حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

        y =

        1    2    3    6

        2    6    3    6

        2    6    6    8

        6    6    6    8

>> mean(y)

        ans =

       2.2500    6.0000    6.0000    6.5000
```

٤. الوسيط Median ويمكن أيجاده بأستخدام الدالة median

يحسب الوسيط بأستخدام حزمة MATLAB وذلك بتطبيق الدالة median ومثال على ذلك نعطي المتغير x صف من أربعة قيم ثم نجد الوسط التوافقي بأستخدام الدالة median وكما يلي:

```
>> x=[1 3 6 2]

    x =

    1    3    6    2

>> median(x)

    ans =

   2.5000
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٤*٤ من القيم ثم نطبق نفس الطريقة لايجاد الوسيط حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

    y =

    1    2    3    6

    2    6    3    6

    2    6    6    8

    6    6    6    8

>> median(y)

    ans =

   2.0000    6.0000    3.5000    7.0000
```

٥. المعدل الحسابي المصحح ويمكن أيجاده بأستخدام الدالة trimmean

يحسب المعدل الحسابي المصحح بأستخدام حزمة MATLAB وذلك بتطبيق الدالة trimmean ومثال على ذلك نعطي المتغير x صف من أربعة قيم ثم نجد الوسط التوافقي بأستخدام الدالة trimmean وتعتمد النتيجة على الحد الثاني في الدالة الذي يحدد النسبة وكما يلي:

```
>> x=[1 3 6 2]

x =

1    3    6    2

>> trimmean(x,10)

ans =

2.5000
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٤*٤ من القيم ثم نطبق نفس الطريقة لايجاد المعدل الحسابي المصحح حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

y =

1    2    3    6

2    6    3    6

2    6    6    8

6    6    6    8

>> trimmean(y,10)

ans =

2.2500    6.0000    6.0000    6.5000
```

```
>> trimmean(y,50)
    ans =
    2.0000    6.0000    3.3333    7.3333
```

٤.٣ مقاييس التشتت Measures of Dispersion

لمقاييس التشتت تطبيقات واسعة في نظم المعلومات حيث أنه بمعرفة مقدار تشتت القيم عن وسطها الحسابي يقودنا الى معرفة معلومات مهمة تستخدم في توجيه البيانات والمعلومات نحو أتجاه معين دون أخر وهذا له أثر كبير في دعم القرارات. لقد تم التطرق في الفصل الأول عن مقاييس التشتت وأستخدامها وأهم المقاييس التي من خلالها تشتت القيم عن وسطها الحسابي هي:

١. قياس المدى ويمكن أيجاده بأستخدام الدالة iqr

يحسب المدى بأستخدام حزمة MATLAB وذلك بتطبيق الدالة iqr ومثال على ذلك نعطي المتغير x صف من ثمانية قيم ثم نجد المدى بأستخدام الدالة iqr وكما يلي:

```
>> x = [2 6 6 8 10 12 16 16]
  x =
  2    6    6    8    10    12    16    16
>> z = iqr(x)
  z =
  8
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٦*٦ قيم ونطبق نفس الطريقة لايجاد المدى حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]
```

```
y =
1   2   3   6
2   6   3   6
2   6   6   8
6   6   6   8
>>w= iqr(y)
w =
1.5000   2.0000   2.0000   3.0000
```

٢. الإنحراف المتوسط المُطلق ويمكن أيجاده بأستخدام الدالة mad

يحسب الإنحراف المتوسط المُطلق بأستخدام حزمة MATLAB وذلك بتطبيق الدالة mad ومثال على ذلك نعطي المتغير x صف من ثمانية قيم ثم نجد المدى بأستخدام الدالة mad

```
>> x = [2 6 6 8 10 12 16 16]
x =
2   6   6   8   10   12   16   16
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٦*٦ قيم ونطبق نفس الطريقة لايجاد الإنحراف المتوسط المُطلق حيث يقوم بحساب كل عمود على أنفراد

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]
y =
1   2   3   6
2   6   3   6
2   6   6   8
6   6   6   8
```

```
>> mad(y)

ans =

0.8750    1.0000    1.0000    1.5000
```

٣. المدى ويمكن أيجاده بأستخدام الدالة range

يحسب المدى بأستخدام حزمة MATLAB وذلك بتطبيق الدالة range ومثال على ذلك نعطي المتغير x صف من ثمانية قيم ثم نجد المدى بأستخدام الدالة range وكما يلي:

```
>> x = [2 6 6 8 10 12 16 16]

x =

2    6    6    8    10    12    16    16

>> range(x)

ans =

16
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٦*٦ قيم ونطبق نفس الطريقة لايجاد المدى حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

y =

1    2    3    6

2    6    3    6

2    6    6    8

6    6    6    8

>> range(y)

ans =

3    6    3    6
```

٤. الانحراف المعيارى ويمكن أيجاده بأستخدام الدالة std

يحسب الانحراف المعيارى بأستخدام حزمة MATLAB وذلك بتطبيق الدالة std ومثال على ذلك نعطي المتغير x صف من ثمانية قيم ثم نجد المدى بأستخدام الدالة std وكما يلي:

```
>> x = [2 6 6 8 10 12 16 16]

   x =

   2    6    6    8   10   12   16   16
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٦*٦ قيم ونطبق نفس الطريقة لايجاد الانحراف المعيارى حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

      y =

      1    2    3    6

      2    6    3    6

      2    6    6    8

      6    6    6    8

>> std(y)

      ans =

     1.2583    1.6330    1.6162    1.9169
```

٥. التباين ويمكن أيجاده بأستخدام الدالة var

يحسب التباين بأستخدام حزمة MATLAB وذلك بتطبيق الدالة var ومثال على ذلك نعطي المتغير x صف من ثمانية قيم ثم نجد التباين بأستخدام الدالة var وكما يلي:

```
>> x = [2 6 6 8 10 12 16 16]

   x =

   2    6    6    8   10   12   16   16
```

```
>> var(x)

ans =

26
```

مثال أخر على ذلك نأخذ متغير اخر y حيث يشير الى مصفوفة ٦*٦ قيم ونطبق نفس الطريقة لايجاد التباين حيث يقوم بحساب كل عمود على أنفراد وكما يلي:

```
>> y=[1 2 3 6;2 6 3 6;2 6 6 8;6 6 6 8]

y =

1    2    3    6

2    6    3    6

2    6    6    8

6    6    6    8

>> var(x1)

ans =

1.5833    2.6667    2.0000    3.6667
```

٤.٤ البياناتِ بالقِيَمِ المفقودةِ وتقديرها

Estimation of Data with Missing Values (NaNs)

في كثيرمن الاحيان يتم فقدان بعض البيانات لاسباب عديدةوهذا يؤثرعلى طبيعة البيانات ودقة تحليلهاويؤدي أيضا الى تحقيق نتائج غيرموثوقة. وعليه فيجب على الباحث أعتمادطرق وأساليب لغرض تقديرهذه البيانات المفقودةوأيجادها.هناك طرق مختلفةلايجادالبيانات المفقودةفي عينةماوهذا له أثركبيرفي التنبأبالنتائج الصحيحة والتي تغذي جميع الانظمةومنهانظم المعلومات ونظم دعم القرارات. كمثال على ذلك يمكن تكوين مصفوفةوحسب ماتعلمناه سابقاوفي هذا المثال نولدمصفوفةبأستخدام الدالة magic

حيث تستخدم هذه الدالة لتوليد مصفوفة مربعة أي أن عدد الصفوف فيها يساوي عدد الاعمدة. ثم بعد ذلك نحدد الصناصر المفقودة ولتكن ١، ٥ ونضع مكانها NaN . ثم في خطوة أخرى بعد ذلك نولد نفس المصفوفة ونحدد العناصر المفقودة فيها ١,٥,٧ وتكتب كما يلي:

```
>> m = magic(3)

m =

    8    1    6

    3    5    7

    6    9    2

>> m([1  5]) = [NaN NaN]

m =

  NaN    1    6

    3  NaN    7

    6    9    2

>> m = magic(3)

m =

    8    1    6

    3    5    7

    6    9    2

>> m([1  5  7]) = [NaN NaN NaN]

m =

  NaN    1  NaN

    3  NaN    7

    6    9    2
```

لتنفيذ عملية حسابية مثلا جمع (sum) قيم الاعمدة في المصفوفة السابقة والتي تَتضمّنُ قيما مفقودةَ والتي تُنتجُ عنها NaN ، فأننا سوف نحصل على نتائج غير صحيحة حيث أن العمودين الاول والثاني كانت نتائجهم خاطئة وكما يلي:

```
>> m = magic(3)

m =

    8    1    6

    3    5    7

    6    9    2

>> m([1 5]) = [NaN NaN]

m =

  NaN    1    6

    3    NaN  7

    6      9    2

>> sum(m)

ans =

  NaN   NaN    15
```

ولحل هذه المشكلة ينفذ الامر أو العملية وكأن الخلايا المحذوفة غير موجودة وتعتبر مهملة في الحسابات وذلك بأستحدام الدالة nansum حيث أن من ميزات حزمة MATLAB أنها تحتوي على دوال بالقيم المفقودة مشابهة للدوال الاعتيادية وكما يلي يحل المثال السابق بأستخدام الدالة nansum وكما يلي:

```
>> nansum(m)

ans =

    7   10   15
```

ومن الدوال ذات القيم المفقودة والمستخدمة في حزمة MATLAB هي:

١. حساب أعلى قيمةُ بالقيم المفقودة nanmax

وتقوم هذه الدالة بحساب القيمة العليا بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعة صفوف وأربعة أعمدة وبأستحدام الدالة magic ثم نطبق حساب أعلى قيمة لكل عمود بأستحدام الدالة max التي تحسب أعلى قيمة لكل عمود في المصفوفة حيث نحصل على النتيجة وبشكل صحيح وتكتب الاوامر كما يلي:

```
>> m=magic(6)

m =

   16    2    3   13
    5   11   10    8
    9    7    6   12
    6   16   15    1

>> max(m)

ans =

   16   16   15   13
```

نكرر ذلك بوضع قيم مفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب أعلى قيمة بأستحدام الدالة max ونجد الفرق بين الحالتين وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]

m =

  NaN      2      3      NaN
    5      11     10     8
    9      7      NaN    12
    6      NaN    15     1
```

```
>> max(m)

ans =

    9   11   15   12
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب أعلى قيمة بأستحدام الدالة nanmax ونجد الفرق بين الحالتين وهذه حالة خاصة لحساب أعلى قيمة لانه هنانختارالقيم الاعلى لوحدها ولاتؤثر على بقية القيم وكما يلي:

```
>> nanmax(m)

ans =

    9   11   15   12
```

٢. حساب المتوسط بالقيم المفقودة nanmean

وتقوم هذه الدالةبحساب قيمةالمتوسط بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعةصفوف وأربعةأعمدة وبأستحدام الدالة magic ثم نطبق حساب قيمةالمتوسط لكل عمود بأستحدام الدالة mean التي تحسب قيمة المتوسط لكل عمود في المصفوفةحيث نحصل على النتيجةوبشكل صحيح وتكتب الاوامر كما يلي:

```
>> m=magic(6)

   m =

   16    2    3   13

    5   11   10    8

    9    7    6   12

    6   16   15    1

>> mean(m)

ans =

   8.5000    8.5000    8.5000    8.5000
```

نكرر ذلك بوضع قيم مفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب قيمة المتوسط بأستحدام الدالة mean حيث نجد أنه هنا كل القيم أصبحت مبهمة وذلك لانه كل الاعمدة في المصفوفة كانت تحتوي على قيم مفقودة وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]
 m =
 NaN  2    3   NaN
   5  11   10   8
   9   7  NaN  12
   6 NaN  15    1
>> mean(m)
ans =
 NaN  NaN  NaN  NaN
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب قيمة المتوسط بأستحدام الدالة nanmean ونجد الفرق بين الحالتين حيث يتم تطبيق العملية مع أهمال القيم المفقودة وكما يلي:

```
>> nanmean(m)
   ans =
  6.0000   6.6667   9.3333   7.0000
```

٣. حساب الوسيط بالقيم المفقودة nanmedian

وتقوم هذه الدالة بحساب قيمة الوسيط بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعة صفوف وأربعة أعمدة وبأستحدام الدالة magic ثم نطبق حساب قيمة الوسيط لكل عمود بأستحدام الدالة median التي تحسب قيمة الوسيط لكل عمود في المصفوفة حيث نحصل على النتيجة وبشكل صحيح وتكتب الاوامر كما يلي:

```
>> m=magic(6)

  m =

  16    2    3   13

   5   11   10    8

   9    7    6   12

   6   16   15    1

>> median(m)

   ans =

   7    9    8   10
```

نكرر ذلك بوضع قيم مفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب قيمة الوسيط بأستحدام الدالة median حيث نجد أنه هنا كل القيم أصبحت مبهمة وذلك لانه كل الاعمدة في المصفوفة كانت تحتوي على قيم مفقودة وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]

m =

  NaN    2     3   NaN

   5    11    10     8

   9     7   NaN    12

   6   NaN    15     1

>> median(m)

ans =

  NaN   NaN   NaN   NaN
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب قيمة الوسيط بأستحدام الدالة nanmedian ونجد الفرق بين الحالتين حيث يتم تطبيق العملية مع أهمال القيم المفقودة وكما يلي:

```
>> nanmedian(m)

ans =

    5    7   10    8
```

٤. حساب أدنى قيمةُ بالقيم المفقودة nanmin

وتقوم هذه الدالة بحساب أقل قيمة بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعة صفوف وأربعة أعمدة وبأستحدام الدالة magic ثم نطبق حساب أقل قيمة لكل عمود بأستحدام الدالة min التي تحسب أقل قيمة لكل عمود في المصفوفة حيث نحصل على النتيجة وبشكل صحيح وتكتب الاوامر كما يلي:

```
>> m=magic(6)

m =

   16    2    3   13
    5   11   10    8
    9    7    6   12
    6   16   15    1

>> min(m)

ans =

    6    2    3    1
```

نكرر ذلك بوضع قيم مفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب أعلى قيمة بأستحدام الدالة min ونجد أنه لا فرق بين الحالتين وذلك لان العملية هنا أختيار العنصر الاقل ولاتؤثر على بقية العناصر وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]

m =

  NaN2      3      NaN

    5       11     10     8

    9       7      NaN    12

    6       NaN    15     1

>> m([1 8 11 13]) = [NaN NaN NaN NaN]

m =

  NaN2      3      NaN

    5       11     10     8

    9       7      NaN    12

    6       NaN    15     1

>> min(m)

ans =

    6    2    3    1
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب أقل قيمة بأستحدام الدالة nanmin ونجد الفرق بين الحالتين وهذه حالة خاصة لحساب أعلى قيمة لانه هنا نختار القيم الاعلى لوحدها ولاتؤثر على بقية القيم وكما يلي:

```
>> nanmin(m)

ans =

    6    2    3    1
```

٥. حساب الأنحراف المعياري بالقيم المفقودة nanstd

وتقوم هذه الدالة بحساب الأنحراف المعياري بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعة صفوف وأربعة أعمدة وبأستحدام الدالة magic ثم نطبق حساب الأنحراف المعياري لكل عمود بأستحدام الدالة std التي تحسب إلأنحراف المعياري لكل عمود في المصفوفة حيث نحصل على النتيجة وبشكل صحيح وتكتب كما يلي:

```
>> m=magic(6)

m =

   16    2    3   13

    5   11   10    8

    9    7    6   12

    6   16   15    1

>> std(m)

ans =

   5.6667    5.1962    5.1962    5.6667
```

نكرر ذلك بوضع قيم مفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب الأنحراف المعياري بأستحدام الدالة std حيث نجد أنه هنا كل القيم أصبحت مبهمة وذلك لانه كل الاعمدة في المصفوفة كانت تحتوي على قيم مفقودة وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]

m =

  NaN2       3       NaN

    5        11      10      8

    9        7       NaN   12

    6        NaN   15      1


>> std(m)

ans =

  NaN   NaN   NaN   NaN
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب إلأنحراف المعياري بأستحدام الدالة nanstd ونجد الفرق بين الحالتين حيث هذه النتيجة تكون الحقيقية وكما يلي:

```
>> nanstd(m)

ans =

   2.6658    6.5092    6.0277    5.5678
```

٦. حساب المجموع بالقيم المفقودة nansum

وتقوم هذه الدالة بحساب المجموع بأهمال القيم المحذوفة ومثال على ذلك ندخل مصفوفة بأربعة صفوف وأربعة أعمدة وبأستحدام الدالة magic ثم نطبق حساب المجموع لكل عمود بأستحدام الدالة sum التي تحسب المجموع لكل عمود في المصفوفة حيث نحصل على النتيجة وبشكل صحيح.

```
>> m=magic(6)
m =
   16        2     3     13
    5        11    10    8
    9        7     6     12
    6        16    15    1
>> sum(m)
ans =
   36   36   36   36
```

نكرر ذلك بوضع قيم مفقودة في عناصرالمصفوفة المرقمة 1,8,11,13 وكما يلي:

```
>> m([1 8 11 13]) = [NaN NaN NaN NaN]
m =
  NaN    2     3     NaN
    5    11    10    8
    9    7     NaN   12
    6    NaN   15    1
```

وبعدها نطبق حساب المجموع بأستحدام الدالة sum حيث نجد أنه هنا كل القيم أصبحت مبهمة وذلك لانه كل الاعمدة في المصفوفة كانت تحتوي على قيم مفقودة وكما يلي:

```
>> sum(m)
ans =
NaN   NaN   NaN   NaN
```

ثم نكرر المصفوفة ذات القيم المفقودة في العناصر 1,8,11,13 وبعدها نطبق حساب المجموع بأستحدام الدالة nansum ونجد الفرق بين الحالتين حيث هذه النتيجة تكون الحقيقية وكما يلي:

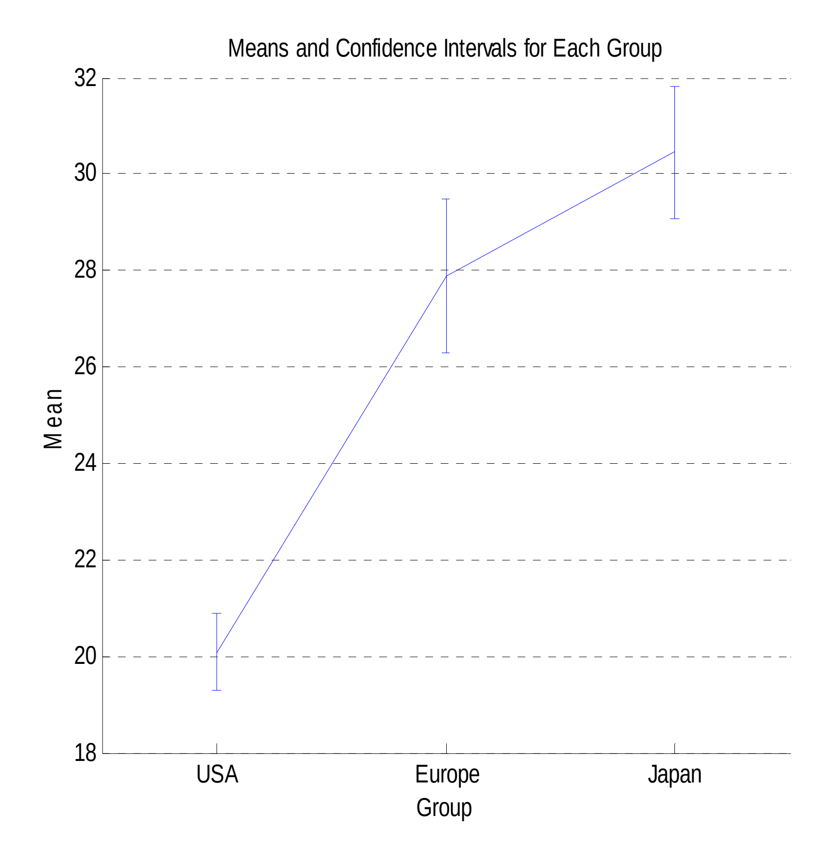
```
>> nansum(m)
   ans =
   18   20   28   21
```

٤.٥ تجميع البياناتِ Grouped Data

عملية تجميع البيانات وربطها مع بعضها لتكوين حقل من البيانات لتوضيح عملية معينة أو لاعداد معلومات واسعة عن موضوع معين كل هذا له أهمية كبيرة في توظيفها في نظم المعلومات ودعم القرارات وهذا يتم بعد دراسة وتحليل هذه البيانات وتوجيهها بالاتجاه الصحيح. تستخدم هذه الطريقة لتجميع البيانات مع بعضها على شكل مجاميع جزئية أو ثانوية (subgroup) حيث يتم التعامل معها بسكل أفضل وأن هذه المجاميع يتم تكوينها وفقا لمواصفات معينة متشابهة أو حسب ما يعتمده الباحث وبعد ذلك يتم حساب الوسط الحسابي والانحراف المعياري لكل مجموعة من تلك المجاميع. فالدالة مثلا grpstats تقوم بحساب الوسط الحسابي لكل عمود في المصفوفة. وفي الَمِثال التالي نحمل الملف carbig والذي يمثل مجموعةَ معلومات عن السيارات ذات الحجم الأكبرِ ويُمْكِنُ أَنْ تحسب قيمة الوسط الحسابي (عدد الأميال لكلّ غالونِ) للسياراتِ والتي جمّعتَ مِن قِبل المنظمة وذلك بأستخدام الدالة grpstats والتي تمثل الصف الاحصائي على شكل مجاميع. والشكل ٤.١ يوضح المجاميع حيث أن القيمة ٢٠.٠٨٣٥ تمثل الولايات المتحدة وأن القيمة ٢٧.٨٩١٦ تمثل المجموعة الاوربية وأن القيمة ٣٠.٦٥٠٦ تمثل اليابان وتكون أوامر التحميل والرسم كما يلي:

```
>> load carbig

>> grpstats(MPG,org,0.05)
```



شكل ٤.١ رسم بياني يمثل ثلاث مجاميع الثانوية للسيارات الكبيرة

يُمْكِنُك أَنْ تَحْصلَ على المجموعةِ الكاملةِ أيضاً مِنْ الإحصائياتِ MPG جمّعتْ بثلاثة متغيّراتِ: المنظمة والمحرّك ذو أربع إسطواناتِ، وصنع السيارة وتكتب الدالة كما يلي:

```
>> [m,s,c,n] = grpstats(MPG,{org cyl6 when});

m =

   16.8961    17.6787    21.5360    23.3333    27.0273    29.7362    17.5000
   30.8333    26.7163    26.9115    35.7000    19.0000    20.8333    26.5000
   26.0833   29.5000   35.3000

s =

   0.3331     0.3023     0.9796     0.8733     0.7566     0.7113     0.9678     3.1761
   0.7308     1.0116     1.6265     0.5776     0.9280     2.0972     1.1772     0.8655
   0.6835

c =

   77    75    25    12    22    38     6     3    21    26    16     3     3     6     12    25     32

n =

   'USA'       'Other'    'Early'
   'USA'       'Other'    'Mid'
   'USA'       'Other'    'Late'
   'USA'       'Four'     'Early'
   'USA'       'Four'     'Mid'
   'USA'       'Four'     'Late'
   'Europe'    'Other'    'Mid'
   'Europe'    'Other'    'Late'
   'Europe'    'Four'     'Early'
   'Europe'    'Four'     'Mid'
   'Europe'    'Four'     'Late'
   'Japan'     'Other'    'Early'
   'Japan'     'Other'    'Mid'
```

```
'Japan'    'Other'    'Late'
'Japan'    'Four'     'Early'
'Japan'    'Four'     'Mid'
'Japan'    'Four'     'Late'
```

وتجمع المعلومات ضمن مصفوفة كما يلي:

```
>> [n num2cell([m s c])]
ans =
 'USA'      'Other'    'Early'    [16.896]    [0.33306]    [77]
 'USA'      'Other'    'Mid'      [17.679]    [0.30225]    [75]
 'USA'      'Other'    'Late'     [21.536]    [0.97961]    [25]
 'USA'      'Four'     'Early'    [23.333]    [0.87328]    [12]
 'USA'      'Four'     'Mid'      [27.027]    [0.75656]    [22]
 'USA'      'Four'     'Late'     [29.736]    [0.71126]    [38]
 'Europe'   'Other'    'Mid'      [  17.5]    [ 0.9678]    [ 6]
 'Europe'   'Other'    'Late'     [30.833]    [ 3.1761]    [ 3]
 'Europe'   'Four'     'Early'    [26.716]    [0.73076]    [21]
 'Europe'   'Four'     'Mid'      [26.912]    [ 1.0116]    [26]
 'Europe'   'Four'     'Late'     [  35.7]    [ 1.6265]    [16]
 'Japan'    'Other'    'Early'    [    19]    [0.57735]    [ 3]
 'Japan'    'Other'    'Mid'      [20.833]    [0.92796]    [ 3]
 'Japan'    'Other'    'Late'     [  26.5]    [ 2.0972]    [ 6]
 'Japan'    'Four'     'Early'    [26.083]    [ 1.1772]    [12]
 'Japan'    'Four'     'Mid'      [  29.5]    [0.86567]    [25]
 'Japan'    'Four'     'Late'     [  35.3]    [0.68366]    [32]
```

٤.٦ وصف البيانات والنسب المؤية

Percentiles and Graphical Descriptions

هنالك طرق مختلفةلوصف البيانات بأستخدام حزمة MATLAB وكل هذه الطرق تؤدي الى فهم أوسع وأدق للبيانات وطرق توزيعها لذا فأن لها أهميةكبيرة في تصميم أنظمة المعلومات وأنظمة أتخاذ القرارات وسوف نتطرق في الفقرات التالية الى عدد من هذه الطرق.

4.6.1 النسبة المئوية Percentiles

وتمثل هذه الطريقة أستخدام النسبة المئوية في وصف البيانات ومثال على ذلك نولد عينة تحتوي على خليط مكون من توزيعين من البيانات بأستخدام الدالة normrnd حيث تستخدم لتوليد مصفوفة برمز x للتوزيع الطبيعي (ملاحظة على ذلك أنه عند تطبيق ذلك من المحتمل أن تكون القيم مختلفة) وكما يلي:

```
>> normrnd(6,1,1,10)

ans =

   3.5676    2.3366    6.1253    6.2877    2.8535

   5.1909    5.1892    3.9626    6.3273    6.1766

>> normrnd(6,0.5,1,20)

ans =

   5.9066    6.3629    5.7058    7.0916    5.9318

   6.0570    6.5336    6.0296    5.9522    5.5838

   6.1672    5.3319    6.3572    6.8118    5.6561

   6.6290    6.6270    5.2031    5.2795    6.2856

>> x = [normrnd(6,1,1,10) normrnd(6,0.5,1,20)]

x =

   3.6001    6.6900    6.8156    6.7119    5.2902
```

```
6.6686    5.1908    2.7975    3.9802    3.8633
5.1980    6.1287    5.6718    6.7076    5.5975
6.2666    6.1097    5.5390    6.9167    5.9706
5.6967    6.3072    6.2539    6.8662    6.2956
5.6782    6.1902    5.6956    5.9902    5.9759
```

بعد ذلك يتم توليد المتغير p لخمسة قيم والتي تمثل النسب المئوية وكما يلي:

```
>> p = 100*(0:0.25:1)
p =
   0    25    50    75   100
```

ثم بعد ذلك نقوم بتكوين المجاميع للعينة وحسب هذه النسب ونرمز لها y وكما يلي:

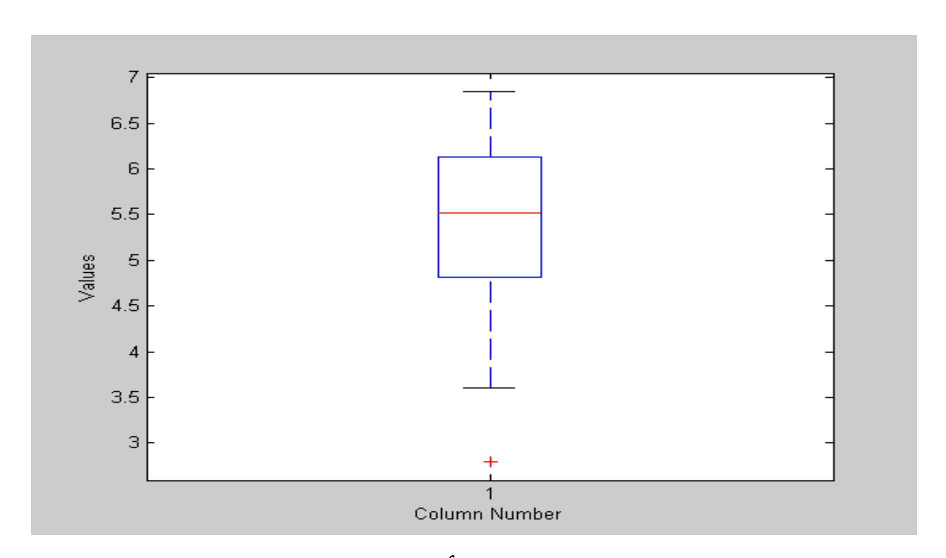
```
>> y = prctile(x,p)
y =
   2.7975    6.8156    5.5172    6.1287    6.8662
```

ثم بعد ذلك نقوم بتكوين المصفوفة لكل من p,y ونرمز لها z وكما يلي

```
>> z = [p;y]
z =
        0        25.0000   50.0000   75.0000  100.0000
   2.7975    6.8156     5.5172     6.1287    6.8662
```

ثم بعد ذلك نقوم برسم العينة باستخدام الدالة boxplot وكما موضحة في الشكل ٤.٢ وكما يلي:

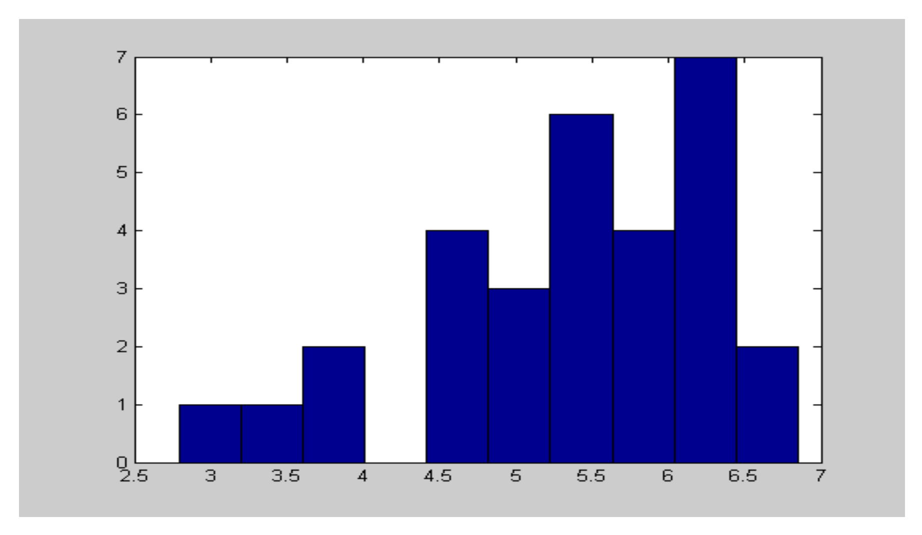
```
>> boxplot(x)
```



شكل ٤.٢ خليط لتوزيعين من البيانات بأستخدام المخطط الصندوقي

وكذلك يمكن تمثيل هذا المثال بأستخدام المخطط المدرج كما في الشكل ٤.٣ بأستخدام الدالة hist وكما يلي:

```
>> hist (x)
```



شكل ٤.٣ خليط لتوزيعين من البيانات بأستخدام المخطط المدرج

4.6.2 تقدير كثافةِ ألاحتمال Probability Density Estimation

تستخدم هذه الطريقة لتخمين وصف كثافة البيانات لعينة معينة من بيئات. ولتطبيق حزمة MATLAB نستخدم الدالة ksdensity والتي تعمل على إيجاد حزمة مرتبطةَ لتَخمين الكثافة.هذا المثالِ يعرض مجموعةَ معلومات السيارةَ الصغيرةَ والمخزونة بالدالة carsmall لتَخمين كثافةِ إحتمالَ قطع عدد أميال لكلّ غالونِ من الوقود MPG المقاييس مطبق بأستخدام ٩٦ سيارةِ. التوزيع يكون طبيعي، وموجته الأصلية موضحة في الشكل ٤.٦.

فالمثال التالي أولا يحمل الملف carsmall ثم يعرض قيم MPG حيث نلاحظ أن بعض القيم مفقودة ثم يعرض المنشأ وبعد ذلك يجمع القيم لتخمين الكثافة بأستخدام الدالة ksdensity وبعدها يرسم الدالة بأستخدام plot وتكتب الاوامر كما يلي:

```
>> cars = load('carsmall','MPG','Origin')

cars =

    Origin: [100x7 char]

    MPG: [100x1 double]

>> MPG = cars.MPG
```

```
MPG =
18.0000   15.0000   18.0000   16.0000   17.0000   15.0000   16.0000
16.0000   16.0000   15.0000   NaN       NaN       NaN       NaN
 NaN      15.0000   16.0000   NaN       15.0000   16.0000   26.0000
22.0000   18.0000   21.0000   27.0000   26.0000   25.0000   26.0000
25.0000   26.0000   21.0000   10.0000   10.0000   11.0000    9.0000
28.0000   25.0000   25.0000   26.0000   27.0000   17.5000   16.0000
15.5000   16.5000   22.0000   22.0000   26.0000   22.5000   29.0000
26.5000   29.0000   33.0000   20.0000   18.0000   18.5000   17.5000
29.5000   32.0000   28.0000   26.5000   20.0000   13.0000   19.0000
19.0000   16.5000   16.5000   13.0000   13.0000   13.0000   28.0000
27.0000   36.0000   31.0000   29.0000   27.0000   26.0000   23.0000
36.0000   37.0000   31.0000   38.0000   36.0000   36.0000   36.0000
36.0000   38.0000   32.0000   38.0000   25.0000   38.0000   26.0000
22.0000   32.0000   36.0000   27.0000   27.0000   66.0000   32.0000
28.0000   31.0000
```

```
>> Origin = cars.Origin

Origin =

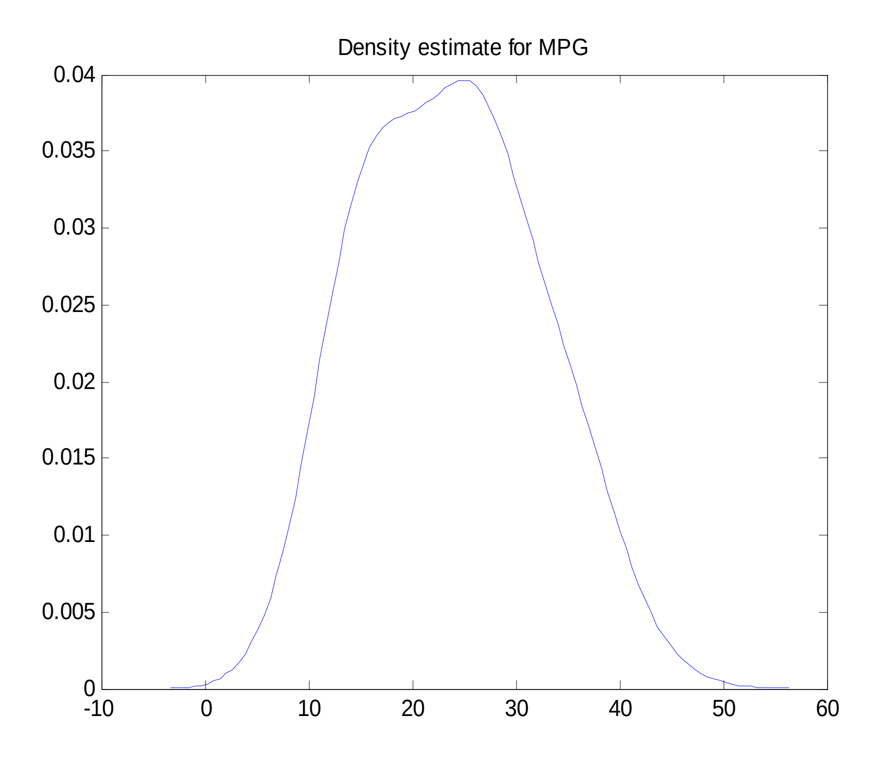
USA     USA     USA     USA     USA     USA     USA     USA     USA     USA
France  USA     USA     USA     USA     USA     USA     USA     USA     USA
Japan   USA     USA     USA     Japan   Germany France  Germany Sweden
Germany USA     USA     USA     USA     USA     Italy   Germany USA     USA
France  USA     USA     USA     USA     USA     USA     USA     USA     USA
USA     Germany Japan   USA     USA     USA     USA     Germany Japan   Japan
USA     Sweden  USA     France  Japan   GermanyUSA      USA     USA     USA
USA     USA     USA     USA     USA     USA     USA     USA     Germany Japan
Japan   USA     USA     Japan   Japan   Japan   Japan   Japan   Japan   USA
USA     USA     USA     Japan   USA     USA     USA     Germany USA     USA     USA
```

```
>> [f,x] = ksdensity(MPG)
f =
    0.0005    0.0006    0.0008    0.0011    0.0016
    0.0018    0.0023    0.0028    0.0035    0.0063
    0.0053    0.0063    0.0076    0.0089    0.0106
    0.0120    0.0138    0.0156    0.0175    0.0195
    0.0215    0.0235    0.0256    0.0272    0.0290
    0.0305    0.0320    0.0332    0.0362    0.0351
    0.0358    0.0363    0.0367    0.0369    0.0372
    0.0373    0.0375    0.0376    0.0378    0.0380
    0.0382    0.0385    0.0388    0.0391    0.0393
    0.0395    0.0396    0.0396    0.0395    0.0392
    0.0388    0.0382    0.0376    0.0365    0.0355
    0.0366    0.0333    0.0321    0.0308    0.0296
    0.0286    0.0272    0.0260    0.0268    0.0237
    0.0225    0.0216    0.0203    0.0192    0.0180
    0.0168    0.0157    0.0165    0.0133    0.0121
    0.0110    0.0098    0.0088    0.0078    0.0068
    0.0059    0.0052    0.0066    0.0038    0.0032
    0.0028    0.0023    0.0020    0.0016    0.0016
    0.0011    0.0009    0.0008    0.0006    0.0005
    0.0006    0.0003    0.0003    0.0002    0.000
```

```
x =
0.7716      1.2912      1.8110      2.3308      2.8505      3.3703
3.8901      6.6098      6.929       5.6696       5.969      6.6889
7.0087      7.5286      8.0682      8.5680      9.0877      9.6075
10.1273     10.6671     11.1668     11.6866     12.2066     12.7261
13.2659     13.7657     16.2856     16.8052     15.3250     15.8667
16.3665     16.8863     17.6060     17.9238     18.6636     18.9636
19.6831     20.0029     20.5227     21.0626      21.5622    22.0820
22.6017     23.1215     23.6613      26.1610    26.6808     25.2006
25.7203     26.2601      26.7599    27.2797     27.7996     28.3192
28.8390     29.3587     29.8785     30.3983     30.9180     31.6378
31.9576     32.6773     32.9971     33.5169     36.0366     36.5566
35.0762     35.5960     36.1157     36.6355     37.1553     37.6750
38.1968     38.7166     39.2363      39.7561    60.2739     60.7936
61.3136     61.8332     62.3529     62.8727     63.3925     63.9123
66.6320     66.9518     65.6716     65.9913     66.5111     67.0309
67.5506     68.0706     68.5902     69.1099     69.6297     50.1695
50.6692     51.1890     51.7088     52.2286
```

```
>> plot(x,f);

>> title('Density estimate for MPG')
```



شكل ٤.٦ التوزيع الطبيعي للسيارات حسب عدد الأميال لكل غالون

أن إختيار موجة Kernal ضمن الدالة يسيطر على تنعيم منحنى كثافةِ الإحتمالَ.فالمثال التالي يوضح تخمينُ الكثافةَ لنفس بيانات المسافة بالأميالِ التي تَستعملُ موجاتَ مختلفةَ. إن الموجة الأصليةَ في والتي تمثل الرسم (*) وتبدو مثل الرسم البياني السَابِقَ. التخمينات للموجات الأصغرِ تمثل الرسم (+) أما التخمينات للموجات الأكبرِ تمثل بالرسم (-) وكما موضحة بالرسم البياني في الشكل ٤.٥ وتكتب الاوامر كما يلي:

```
>> [f,x,u] = ksdensity(MPG)
```

```
f =
0.0005    0.0006    0.0008    0.0011    0.0016    0.0018
0.0023    0.0028    0.0035    0.0063    0.0053    0.0063
0.0076    0.0089    0.0106    0.0120    0.0138    0.0156
0.0175    0.0195    0.0215    0.0235    0.0256    0.0272
0.0290    0.0305    0.0320    0.0332    0.0362    0.0351
0.0358    0.0363    0.0367    0.0369    0.0372    0.0373
0.0375    0.0376    0.0378    0.0380    0.0382    0.0385
0.0388    0.0391    0.0393    0.0395    0.0396    0.0396
0.0395    0.0392    0.0388    0.0382    0.0376    0.0365
0.0355    0.0366    0.0333    0.0321    0.0308    0.0296
0.0286    0.0272    0.0260    0.0268    0.0237    0.0225
0.0216    0.0203    0.0192    0.0180    0.0168    0.0157
0.0165    0.0133    0.0121    0.0110    0.0098    0.0088
0.0078    0.0068    0.0059    0.0052    0.0066    0.0038
0.0032    0.0028    0.0023    0.0020    0.0016    0.0016
0.0011    0.0009    0.0008    0.0006    0.0005    0.0006
0.0003    0.0003 0.0002    0.0002
```

```
x =
0.7716    1.2912    1.8110    2.3308    2.8505    3.3703
3.8901    6.6098    6.9296    5.6696    5.9691    6.6889
7.0087    7.5286    8.0682    8.5680    9.0877    9.6075
10.1273   10.6671   11.1668   11.6866   12.2066   12.7261
3.2659    13.7657   16.2856   16.8052   15.3250   15.8667
16.3665   16.8863   17.6060   17.9238   18.6636   18.9636
19.6831   20.0029   20.5227   21.0626   21.5622   22.0820
22.6017   23.1215   23.6613   26.1610   26.6808   25.2006
25.7203   26.2601   26.7599   27.2797   27.7996   28.3192
28.8390   29.3587   29.8785   30.3983   30.9180   31.6378
31.9576   32.6773   32.9971   33.5169   36.0366   36.5566
35.0762   35.5960   36.1157   36.6355    37.1553  37.6750
38.1968   38.7166   39.2363   39.7561   60.2739   60.7936
61.3136   61.8332   62.3529   62.8727   63.3925   63.9123
66.6320   66.9518   65.6716   65.9913   66.5111   67.0309
67.5506   68.0706   68.5902   69.1099   69.6297   50.1695
50.6692   51.1890   51.7088   52.2286
```

```
u =

        6.1163


>> plot(x,f,'*')

>> title('Density estimate for MPG')

>> hold on


>> [f,x] = ksdensity(MPG,'width',u/3)
```

```
f =
0.0006    0.0011    0.0018    0.0029    0.0066    0.0061    0.0079
0.0096    0.0110    0.0122    0.0131    0.0161    0.0157    0.0181
0.0217    0.0263    0.0317    0.0372    0.0622    0.0660    0.0686
0.0692    0.0688    0.0675    0.0659    0.0666    0.0630    0.0617
0.0602    0.0386    0.0362    0.0337    0.0316    0.0296    0.0285
0.0283    0.0288    0.0299    0.0312    0.0327    0.0366    0.0365
0.0390    0.0618    0.0667    0.0675    0.0500    0.0519    0.0530
0.0533    0.0526    0.0509    0.0683    0.0668    0.0608    0.0368
0.0331    0.0301    0.0281    0.0271    0.0268    0.0267    0.0265
0.0258    0.0266    0.0229    0.0213    0.0200    0.0195    0.0197
0.0207    0.0222    0.0238    0.0250    0.0255    0.0252    0.0260
0.0219    0.0190    0.0157    0.0122    0.0089    0.0060    0.0038
0.0023    0.0015    0.0011    0.0011    0.0015    0.0020    0.0025
0.0029    0.0031    0.0030    0.0027    0.0022    0.0017    0.0012
0.0007    0.0006
```

```
x =
6.2571    6.6661    7.0750    7.6860    7.8929    8.3019    8.7108
9.1198    9.5287    9.9377    10.3666   10.7556   11.1665   11.5735
11.9826   12.3913   12.8003   13.2092   13.6182   16.0271   16.6361
16.8650   15.2560   15.6629   16.0719   16.6808   16.8898   17.2987
17.7077   18.1166   18.5255   18.9365   19.3636   19.7526   20.1613
20.5703   20.9792   21.3882   21.7971   22.2061   22.6150   23.0260
23.6329   23.8618   26.2508   26.6597   25.0687   25.6776   25.8866
26.2955   26.7065   27.1136   27.5226   27.9313   28.3603   28.7692
29.1582   29.5671   29.9760   30.3850   ٣0.7939   31.2029   31.6118
32.0208   32.6297   32.8387   33.2676   33.6566   36.0655   36.6765
36.8836   35.2923   35.7013   36.1102   36.5192   36.9281   37.3371
37.7660   38.1550   38.5639   38.9729   39.3818   39.7908   60.1997
60.6087   61.0176   61.6265   61.8355   62.2666   62.6536   63.0623
63.6713   63.8802   66.2892   66.6981   65.1071   65.5160   65.9250
66.3339   66.7629
```

```
>> plot(x,f,'+')


>> [f,x] = ksdensity(MPG,'width',u*3)

f =

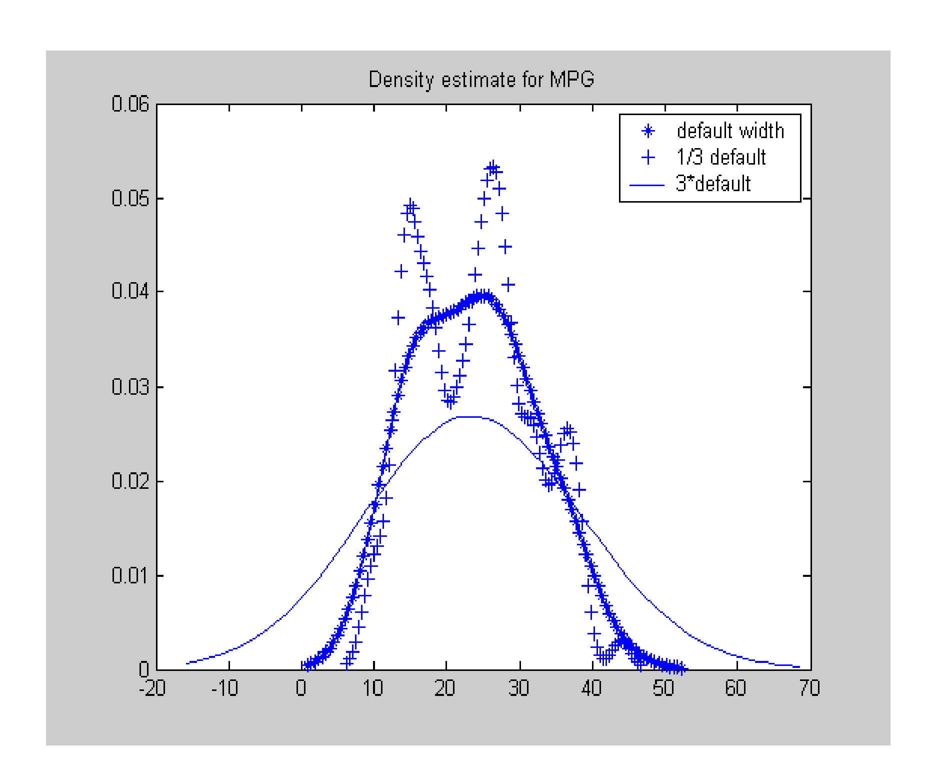
    0.0007    0.0008    0.0009    0.0011    0.0013    0.0015    0.0017    0.0020
    0.0023    0.0026    0.0029    0.0033    0.0038    0.0062    0.0068    0.0053
    0.0059    0.0066    0.0073    0.0080    0.0087    0.0095    0.0106    0.0112
    0.0121    0.0131    0.0160    0.0169    0.0159    0.0168    0.0178    0.0187
    0.0196    0.0205    0.0216    0.0222    0.0229    0.0237    0.0263    0.0269
    0.0256    0.0259    0.0262    0.0265    0.0267    0.0268    0.0268    0.0268
    0.0266    0.0266    0.0261    0.0257    0.0253    0.0267    0.0261    0.0235
    0.0228    0.0220    0.0212    0.0206    0.0195    0.0186    0.0177    0.0168
    0.0159    0.0150    0.0161    0.0132    0.0123    0.0116    0.0106    0.0098
    0.0090    0.0083    0.0076    0.0069    0.0063    0.0057    0.0051    0.0066
    0.0061    0.0037    0.0033    0.0029    0.0026    0.0022    0.0020    0.0017
    0.0015    0.0013    0.0011    0.0010    0.0008    0.0007    0.0006    0.0005
    0.0006    0.0006    0.0003    0.0003
```

```
x =
-15.6857   -16.8336   -13.9812   -13.1290   -12.2767
-11.6265   -10.5723   -9.7200    -8.8678    -8.0155
-7.1633    -6.3111    -5.6588    -6.6066    -3.7566
-2.9021    -2.0699    -1.1977    -0.3656    0.5068
1.3590     2.2113     3.0635     3.9158     6.7680
5.6202     6.6725     7.3267     8.1769     9.0292
9.8816     10.7336    11.5859    12.6381    13.2903
16.1626    16.9968    15.8671    16.6993    17.5515
18.6038    19.2560    20.1082    20.9605    21.8127
22.6669    23.5172    26.3696    25.2216    26.0739
26.9261    27.7786    28.6306    29.6828    30.3351
31.1873    32.0395    32.8918    33.7660    36.5962
35.6685    36.3007    37.1529    38.0052    38.8576
39.7097    60.5619    61.6161    62.2666    63.1186
63.9708    66.8231    65.6753    66.5275    67.3798
68.2320    69.0862    69.9365    50.7887    51.6610
52.6932    53.3656    56.1977    55.0699    55.9021
56.7566    57.6066    58.6588    59.3111    60.1633
61.0155    61.8678    62.7200    63.5723    66.6265
65.2767    66.1290    66.9812    67.8336    68.6857
```

```
>> plot(x,f,'-')

>> legend('default width','1/3 default','3*default')

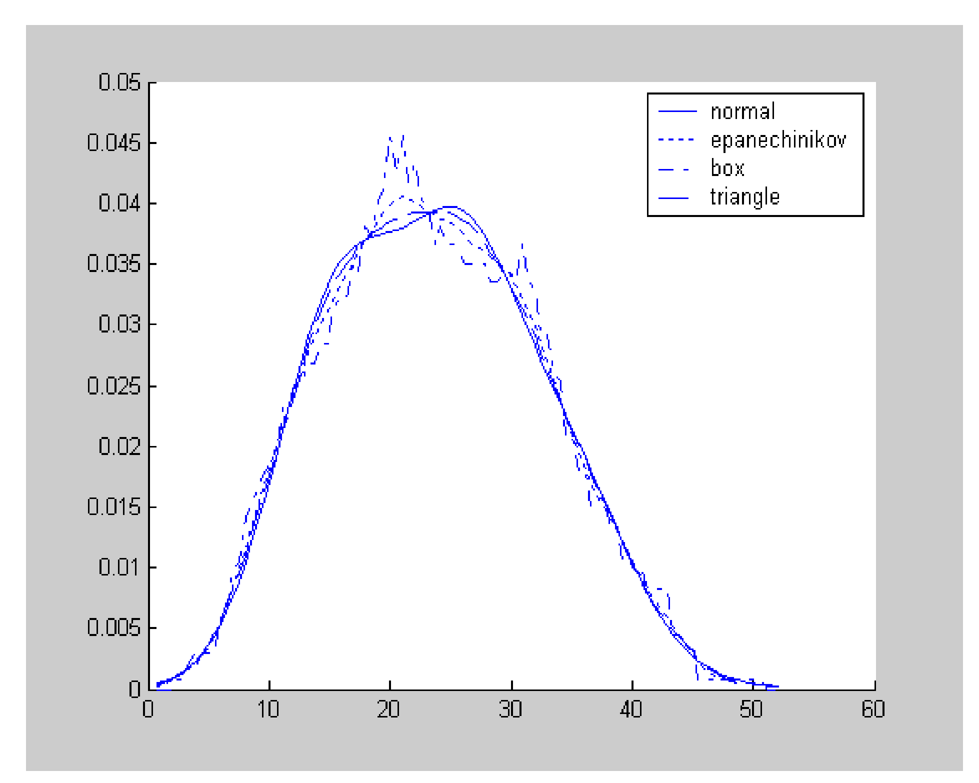
>> hold off
```



شكل ٤.٥ تخمين الكثافة لنفس البيانات السابقة ولموجات مختلفة وحسب الوزن

إستعمال الموجاتِ الأصليةِ، يُمْكِنُك أَنْ تُرسم نفس البياناتِالخاصة بالسيارات الصغيرة حسب المسافة بالأميالِ لكل غالون من الوقود ففي المثال التالي نحمل الملف الخاص بالبيانات بأستخدام الدالة load ثم نؤشر على كل دالة 'وضع طبيعي (normal)' ، ' (epanechinikov)، 'صندوق (box)'، و'مثلث (triangle)' وبعدها نستخدم الدالة ksdensity لحساب تحمين الكثافة ثم بعد ذلك نرسم المتغيرات للمنحنيات الاربعة وكما في الشكل ٤.٦ وتكتب الاوامر كما يلي:

```
>> Load carsmall;
>> hname = {'normal' 'epanechinikov' 'box' 'triangle'};
>> hold on;
>> colors = {'-' ':' '-.' '--'};
>> for j=1:6
>> [f,x] = ksdensity(MPG,'kernel',hname{j});
>> plot(x,f,colors{j});
>> end
>> legend(hname{:});
>> hold off
```



شكل ٤.٦ تخمينُ الكثافةَ لنفس البياناتِ السابقة ولأربع دوال مختلفة

4.6.3 التوزيعِ التراكمي التجريبي

Empirical Cumulative Distribution

يستخدم التوزيع التراكمي التجريبي لانتاج نسخة تجريبية من دالة التوزيع التراكمي cdf ضمن العينة المختارة. أما الدالة ecdf فهي دالة التوزيع التراكمي التخمينية أو التجريبية أي يُمْكِنك أَنْ تَستعملَ ecdf لحِساب cdf التجريبي والدرجات التَخطيطه ففي المثالُ التالي وكما في الشكل ٤.٧ ويمكن توضيح خطوات المثال كما يلي:

١. توليد ٢٠ قيمةُ مِنْ التوزيع الطبيعي بمتوسطِ ١٠ وإنحراف معياري ٢ بأستخدام الدالة normrnd .

٢. حساب القيم التجريبية لهذه العينة بأستخدام الدالة ecdf.

٣. ثمّ نرسم المنحني التدريجى بأستحدام الدالة stairs وهو المنحني المتصل في الشكل.

٤. نولد مئة تدريجة بأستخدام الدالة linspace .

٥. ثم نجد دالة التركم الطبيعي cdf بأستحدام الدالة normcdf .

٦. نستخدم الدالة plot لرسم البيانات وهو المنحني وتكتب الاوامر كما يلي:

```
>> x = normrnd(10,2,20,1)
```

```
x =
11.2665   11.5981   11.8818   8.0158   10.6261   10.6758
7.9865    8.5159    12.1666   9.7370   10.7798   10.1760
8.7291    8.8809    10.8873   8.1002   11.5626   11.1379
8.3566    9.6688
```

```
>> [f,xf] = ecdf(x)
```

```
f =
0.0000    0.0500    0.1000    0.1500    0.2000
0.2500    0.3000    0.3500    0.6000    0.6500
0.5000    0.5500    0.6000    0.6500    0.7000
0.7500    0.8000    0.8500    0.9000    0.9500
1.0000
```

```
xf =
7.9865    7.9865    8.0158    8.1002    8.3566
8.5159    8.7291    8.8809    9.6688    9.7370
10.1760   10.6261   10.6758   10.7798   10.8873
1.1379    11.2665   11.5626   11.5981   11.8818
12.1666
```

```
>> stairs(xf,f)
```

```
>> xx=linspace(5,15,100)
```

```
xx =
5.0000     5.1010     5.2020     5.3030     5.6060     5.5051
5.6061     5.7071     5.8081     5.9091     6.0101     6.1111
6.2121     6.3131     6.6161     6.5152     6.6162     6.7172
6.8182     6.9192     7.0202     7.1212     7.2222     7.3232
7.6262     7.5253     7.6263     7.7273     7.8283     7.9293
8.0303     8.1313     8.2323     8.3333     8.6363     8.5356
8.6366     8.7376     8.8386     8.9396     9.0606     9.1616
9.2626     9.3636     9.6666     9.5655     9.6665     9.7675
9.8685     9.9695     10.0505    10.1515    10.2525    10.3535
10.6565    10.5556    10.6566    10.7576    10.8586    10.9596
11.0606    11.1616    11.2626    11.3636    11.6666    11.5657
11.6667    11.7677    11.8687    11.9697    12.0707    12.1717
12.2727    12.3737    12.6767    12.5758    12.6768    12.7778
12.8788    12.9798    13.0808    13.1818    13.2828    13.3838
13.6868    13.5859    13.6869    13.7879    13.8889    13.9899
16.0909    16.1919    16.2929    16.3939    16.6969    16.5960
16.6970    16.7980    16.8990    15.0000
```

```
>> yy = normcdf(xx,10,2)
```

```
yy =

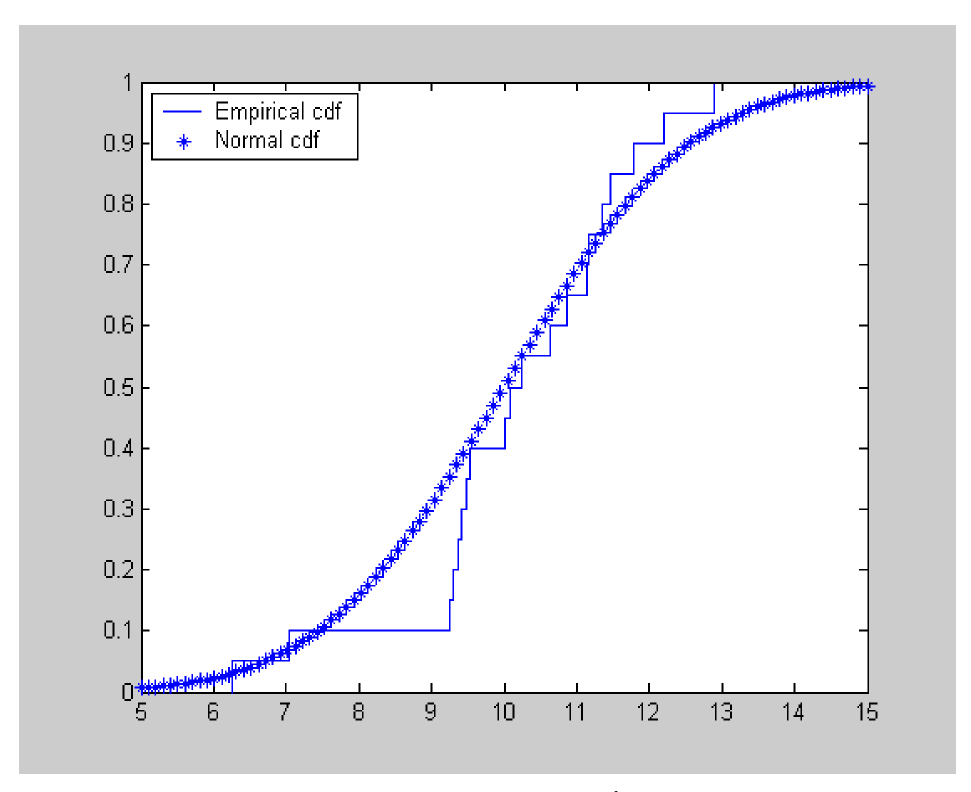
0.0062   0.0072   0.0082   0.0096   0.0108   0.0123   0.0160   0.0159
0.0180   0.0206   0.0230   0.0259   0.0291   0.0326   0.0365   0.0607
0.0653   0.0506   0.0558   0.0617   0.0681   0.0750   0.0826   0.0906
0.0989   0.1080   0.1176   0.1279   0.1388   0.1503   0.1623   0.1751
0.1886   0.2023   0.2169   0.2320   0.2677   0.2639   0.2807   0.2980
0.3157   0.3339   0.3526   0.3713   0.3906   0.6101   0.6298   0.6698
0.6698   0.6899   0.5101   0.5302   0.5502   0.5702   0.5899   0.6096
0.6287   0.6676   0.6661   0.6863   0.7020   0.7193   0.7361   0.7523
0.7680   0.7831   0.7977   0.8116   0.8269   0.8377   0.8697   0.8612
0.8721   0.8826   0.8920   0.9011   0.9096   0.9176   0.9250   0.9319
0.9383   0.9662   0.9696   0.9567   0.9593   0.9635   0.9676   0.9709
0.9761   0.9770   0.9796   0.9820   0.9861   0.9860   0.9877   0.9892
0.9906   0.9918   0.9928   0.9938
```

```
>> hold on;

>> plot(xx,yy,'*')

>> hold off

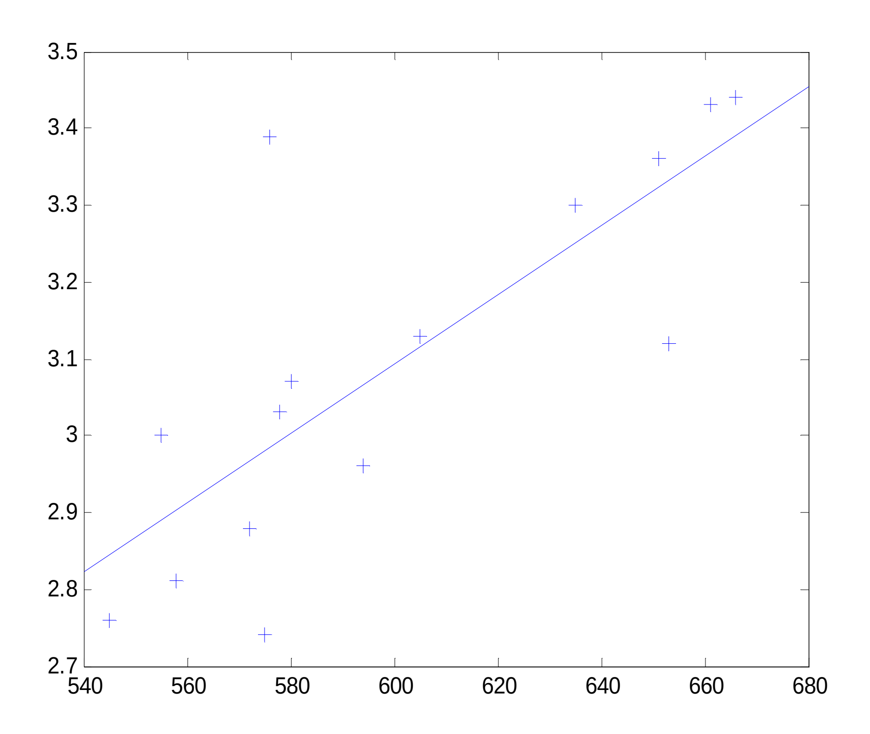
>> legend('Empirical cdf','Normal cdf',2)
```



شكل 4.7 مثال يوضح ٢٠ ملاحظةُ مِنْ توزيع طبيعي بمتوسطِ ١٠ وإنحراف معياري ٢

المثال التالي هو مثال حقيقي أَخذ من مؤسسة Efron و Tibshirani لعام 1993حيث يُقارن كيفية إختبارَ دخولِ الطلبة الى كليةِ حقوق (LAST) فهناك أعداد كبيرة ومعدل نقطة درجة كليةِ الحقوق (GPA) فالعيّنة مكونة مِن ١٥ كليةِ حقوق ففي البداية نحمل المعلومات من الملف lawdata وهي مصنفة ثم نرسمها بأستخدام الدالة plot وتكون ذات الاشارة (+) ثم بعد ذلك نستخدم الدالة lsline والتي تعمل على رسم خط مستقيم لكل البيانات المشتتة وكما موضح في الشكل (٤.٨) وهو يبين أن أعداد كبيرة من الطلبة في LAST سوف تذهب الى كليات الحقوق الاعلى معدل GPA وتكتب الاوامر كما يلي:

```
>> load lawdata
>> plot(lsat,gpa,'+')
>> lsline
```



شكل ٤.٨ عينة كليات الحقوق والقبول فيها لعام ١٩٩٣

الان بألأمكان أَنْ تحسب معامل إرتباط المتغيرات السابقة LAST,GPA بأَستعمال الدالة corrcoef وتكتب الاوامر كما يلي:

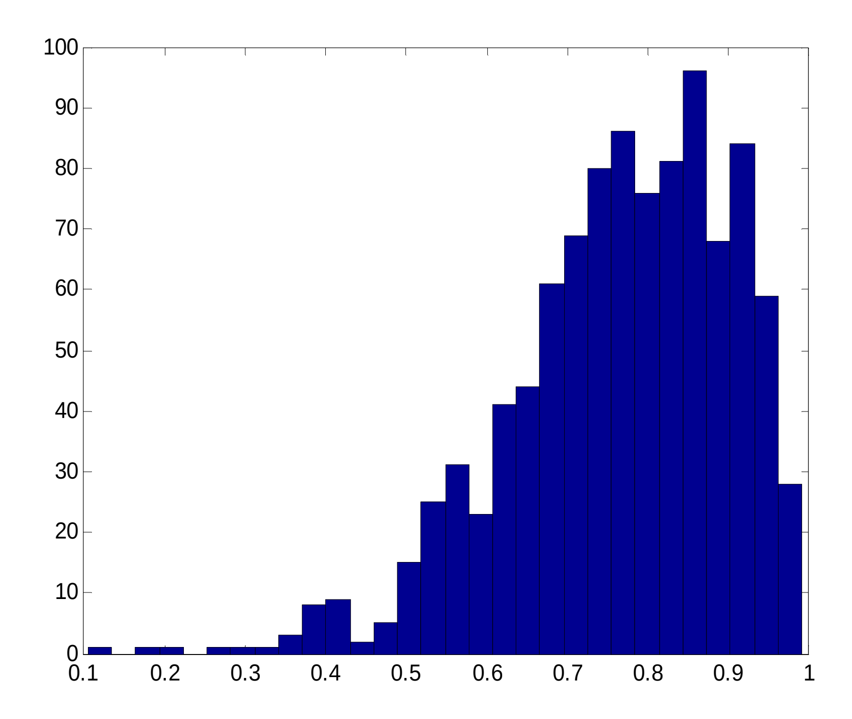
```
>> rhohat = corrcoef(lsat,gpa)
>> rhohat =
   1.0000    0.7766
   0.7766    1.0000
```

النتيجة لديك العدد ٠.٧٧٦٦   والذي يَصِف الإتّصالَ الإيجابيَ بين LAST و GPA، لكن مع ذلك ٠.٧٧٦٦ قَدْ يَبْدو كبير.

ويمكن إستعمال الدالة bootstrp حيث يمكن إعادة عيّنةِ lsat وموجهات gpa عدة مرات حسب ما تريد وتعتمد الإختلافَ في معاملاتِ الإرتباطِ الناتجةِ وفي المثال التالي فأن موجهات last , gpa مكونة من ١٠٠٠ قيمة ويَحْسبانِ الدالة corrcoef على كُلّ عيّنة والشكل ٤.٩ يمثل المدرج الاحصائي بأستخدام الدالة hist لهذه النتيجةِ وتكتب الاوامر كما يلي:

```
>> rhos1000 = bootstrp(1000,'corrcoef',lsat,gpa);

>> hist(rhos1000(:,2),30)
```



شكل ٤.٩ المدرج الإحصائي للعينة في المثال السابق مكونة من ١٠٠٠ قيمة

الفصل الخامس

تطبيقات حزمة MATLAB

في الرسوم

والمخططات الإحصائية

Matlab Applications

in Statistical Plots and Graphics

5.1 مقدمة **Introduction**

٥.٢ مخططات الصندوقِ **Box plots**

5.3 مخططات التوزيعِ **Distribution Plots**

٥.٣.١ مخططات الإحتمالِ الطبيعي **Normal Probability Plots**

٢.٣.٥ أختبار توزيع عينتين بأستخدام دالة الرسم

Quantile-quantile plots

٥.٣.٣ أختبار توزيع عينتين من مجتمعين بأستخدام دالة الرسم

Weibull Probability Plots

٥.٤ التوزيعِ التجميعي

Empirical Cumulative Distribution Function (CDF)

٥.٥ رسومات التشتت **Scatter Plots**

٥.١ مقدمة Introduction

البيانات المعروضة على شكل مخططات ومنحنيات لها تأثير كبير على المستخدم حيث بأمكانه أن يستنبط منها معلومات مهمة وفعالة بعكس البيانات الموجودة على شكل قيم وأرقام فأنها تحتاج الى تحليل ومعالجة لكي يتم فهمها وأستخلاص المهم منها. تعتبر الطرق المستخدمة في عرض البيانات على شكل مخططات ورسومات لها أهمية كبيرة في توضيح وأدراك واسع للعينات وتحديد مدى التقارب أو الاختلاف بين العينات وهذا يتم بشكل دقيق وفعال بأستخدام هذه الطرق حيث يلاحظ لها أهمية كبيرة في تطبيقات نظم المعلومات ونظم أتخاذ القرارات.

تعتبر تطبيقات الرسومات والمخططات ذوات أهمية كبيرة من حيث استخداماته الواسعة في مجالات مختلفة. تُضيف البرامج الإحصائية في حزمة MATLAB مخططات متخصصة ولها قابليات الرسومات الشاملة بأستخدام حزمة MATLABومن هذه الرسومات والمخططات نستعرض التالي:

١. مخططات الصندوقِ (boxplot) وهي رسوم بيانية لوَصْف عيناتِ البياناتِ. وأيضاً ذات فائدة للمقارناتِ التخطيطيةِ بين المتوسط لعدد مِنْ العيناتِ (تَرى ذلك في التحليلَ أحادي الإتجاهَ للتباينِ (ANOVA)).

٢. مخططات التوزيعِ (distribution plots) وهي رسوم بيانية لتصوير وتوضيح توزيعِ عينة واحدة أَو أكثرِ. ويتضمن ذلك مخططات (normal and Weibull probability) ومخططات (quantile-quantile) ومخططات (empirical cumulative distribution) .

٣. مخططات التشتت (scatter plots) وهي رسوم بيانية لتصوير وتوضيح العلاقةِ بين زوج أَو عِدّة أزواجِ من المتغيّراتِ.

وسوف نستعرض هذه الانواع بالتفصيل مع أمثلة تطبيقية عليها.

٥.٢ مخططات الصندوقِ Box plots

يستخدم المخطط الصندوقي للدلالة على العينات بأستخدام الرسم باتجاهين. المثال التالي يوضح عينة من تسعة قيم حيث تزداد تدريجيا ففي البداية نولد القيم أي نعرف المتجه المتغير x حيث يمثل تسعة عناصر وكما يلي:

```
>> x = [2 4 6 8 10 12 14 16 18]

   x =

   2    4    6    8   10   12   14   16   18
```

ثم نرسم المتغير x كما في الشكل ٥.١ باستخدام الدالة boxplot وكما يلي:

```
>> boxplot(x)
```

يمكن أعادة المثال السابق باستخدام المصفوفة y بثلاث صفوف وثلاث أعمدة ففي البداية نعرف المصفوفة بقيمها التسعة وكما يلي:

```
>> y = [2 4 6; 8 10 12; 14 16 18]

y =

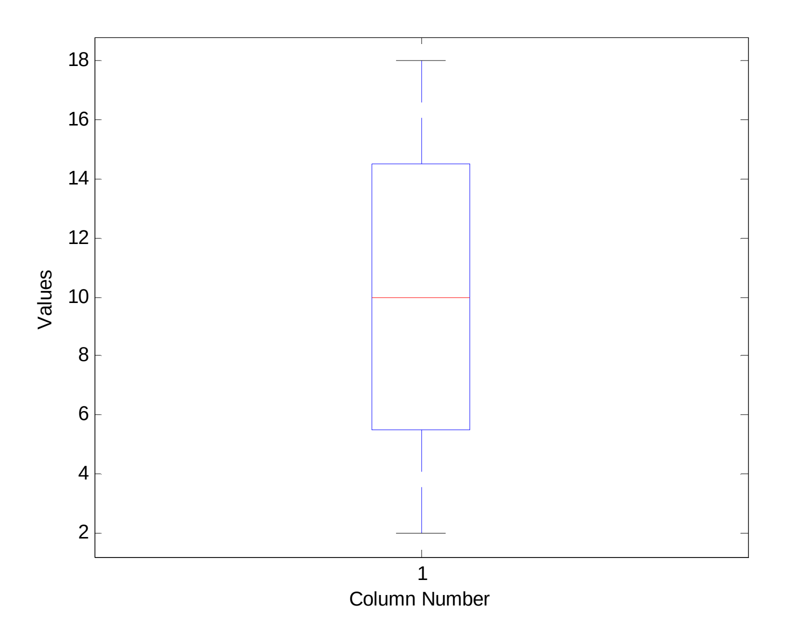
   2    4    6

   8   10   12

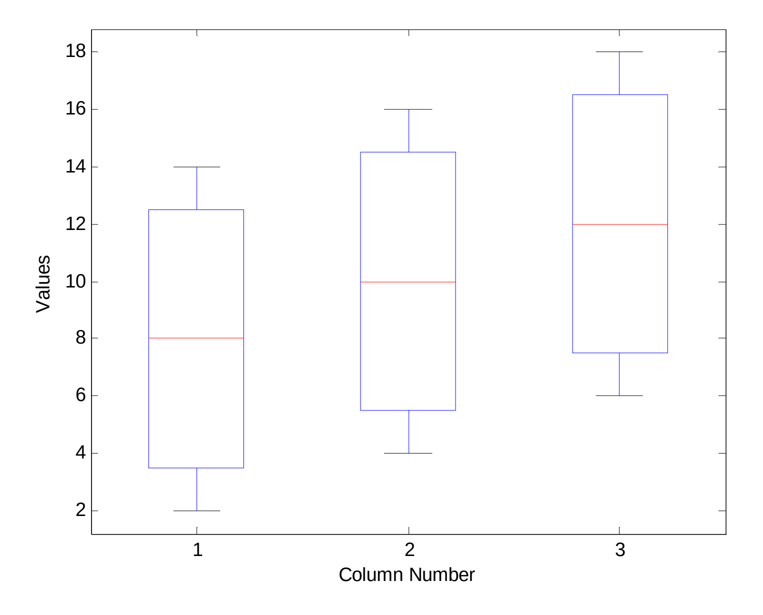
  14   16   18
```

ثم نرسم المتغير y كما في الشكل ٥.٢ باستخدام الدالة boxplot وكما يلي:

```
>> boxplot(y)
```



شكل ٥.١ مخطط يمثل عينة من تسعة قيم حيث تزداد تدريجيا



شكل ٥.٢ مخطط يمثل عينة لمصفوفة ٣*٣

٥.٣ مخططات التوزيعِ Distribution Plots

هناك عِدّة أنواع من المخططات لفَحْص توزيعِ عينة واحدة أَو أكثر من العيناتِ ، وسوف نستعرضها مع الامثلة التطبيقية لكل نوع منها.

٥.٣.١ مخططات الإحتمالِ الطبيعي Normal Probability Plots

مخطط الاحتمالِ الطبيعي عبارة عن رسم بياني مفيد للتَقييم أذا كانت البياناتِ تأتي مِنْ توزيع طبيعي. أن العديد مِنْ الإجراءاتِ الإحصائيةِ تساند الفرضيةِ التي تقول أن النقاط التي يكون انتشارها تحت خط التوزيعِ للبياناتِ يكون طبيعيُ، لذا فأن هذا المخطط يُمْكِنُ أَنْ يستفاد منه في التحقق من هذه الفرضيةَ.

المثال التالي يوضح أستخدام الدالة normrnd وهي تمثل الصفوف العشوائية مِنْ التوزيعِ الطبيعيِ أي Random arrays from the normal distribution حيث تم أخذ معدل القيم ١٠ والانحراف المعياري ١ ولخمس وعشرون قيمة. بعدها تم رسمت البيانات باستخدام الدالة normplot وكما موضحة في الشكل ٥.٣. لدراسة المخطط في الشكل ٥.٣ نلاحظ أن لَه ثلاثة أجزاء أو عناصر أساسية تشمل المنحني المرسوم بالإشارة زائد (+) يوضح الاحتمال التجريبيَ مقابل قيم البياناتِ لكُلّ نقطة في العيّنةِ أما المنحني المرسوم بإشارة الخَطُّ المتصل يمثل النسبة المئوية للخامس والعشرون والخامس والسبعون للبياناتِ وهي عبارة عن علاقة خطيّة. أما المنحني المرسوم بإشارة الخَطُّ المتقطع يكمل الخَطُّ المتصل إلى نهاياتِ العيّنةِ وهو كذلك عبارة عن علاقة خطيّة. أن قياس الإحداثي الصادي لَيسَ موحّدَ وتمثل قِيَمَ الإحداثي الصادي بعناصر الاحتمالات، وهو يتغير مِنْ صفر إلى واحد وأن المسافات بين التقسيمات على الإحداثي الصادي تتطابق مع المسافات في التوزيع الطبيعي. أما المحور السيني فيمثل بعنصر البيانات وتكون التقسيمات متساوية وتتغير من ٨ الى ١٢.

لتوليد العينة نفرض أن معدل القيم هو μ وأن الانحراف المعياري هو σ وأن حجم العينة هو N وأنها موجودة في عمود واحد هو C وعليه يمكن كتابة الدالة كما يلي:

```
x = normrnd(μ,σ ,N,C)
```

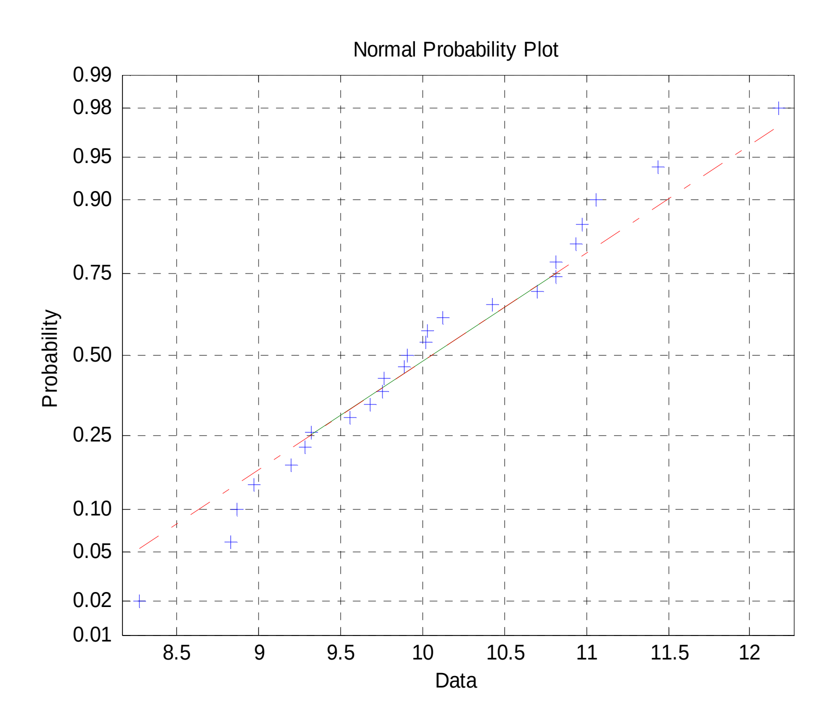
وباختيار μ=١٠ وσ=١ و N=٢٥ و C=١ يمكن أن نكتب الدالة normrnd على الشاشة الرئيسية كما يلي :

```
>> x = normrnd(10,1,25,1)
```

```
x =
12.1764   10.4316   9.5562   10.0300   9.6843
10.9778   10.0183   10.8180  10.7023   9.7687
9.8863    10.1279   9.2006   9.7614    9.9105
8.9767    10.9375   8.8683   9.2893    8.8305
11.0654   9.3196    8.2742   10.8132   11.4419
```

ثم بعد ذلك نستخدم الدالة normplot لرسم العينة x وكما يلي:

```
>> normplot(x)
```



شكل ٥.٣ رسم البيانات ذات الصفوف العشوائية مِنْ التوزيعِ الطبيعيِ بأستخدام normplot

إذا كانت كل نقاط البياناتَ تقع قُرْب الخَطِّ، فأن فرضية الحالة الطبيعيةِ تكون معقولةُ لكن إذا كانت البياناتِ في وضع غير طبيعي، فأن توزيع البيانات ذات إشارات زائدَ (+) قَدْ تتبع المنحنى، كما في المثالِ التالي والذي تستعملُ فيه بياناتَ أسّيةَ.

المثال التالي يوضح رسم الدالة الأسيةأستخدام الدالة exprnd وهي تمثل الصفوف العشوائية مِنْ التوزيعِ الأسّيِ أي Random arrays from exponential distribution حيث تم أخذ معدل القيم ١٠ والانحراف المعياري ١ ولخمس وعشرون قيمة. بعدها تم رسم البيانات بأستخدام الدالة normplot وكما موضحة في الشكل ٥.٤ حيث يكون مشابه للشكل السابق ولكن الاختلاف أن البيانات موزعة بشكل أسي. أن مِقياس الإحداثي الصادي لَيسَ موحّد َوتمثل قِيَمَ الإحداثي الصادي بعنصر الإحتمالاتَ، وهو يتغير مِنْ صفر

إلى واحد. أما المحور السيني فيمثل بعنصر البيانات وتكون التقسيمات متساوية وتتغير من ٠ الى ٤٠.

لتوليد العينة نفرض أن معدل القيم هو μ وأن حجم العينة هو N وأنها موجودة في عمود واحد هو C وعليه يمكن كتابة الدالة كما يلي:

```
x = exprnd(μ, N,C)
```

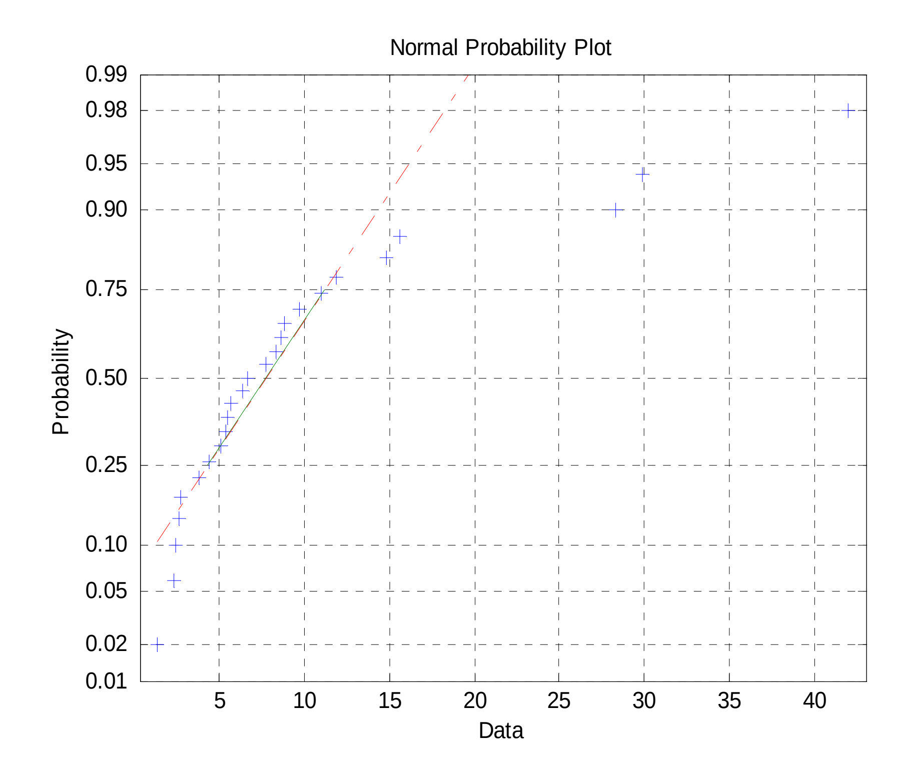
وباختيار ١٠=μ و ٢٥=N و ١=C يمكن أن نكتب الدالة exprnd على الشاشة الرئيسية كما يلي :

```
>> x = exprnd(10,25,1)
```

```
x =
0.2959     0.0997     2.3716     8.2403    6.9653     15.4195
4.4085     11.3932    0.4072     3.1934    8.8685     2.9495
13.1697    8.2115     0.6894     3.8077    15.4853    1.7526
4.6397     20.1161    15.7440    4.9890    4.6221     9.9296
5.5313
```

ثم بعد ذلك نستخدم الدالة normplot لرسم العينة x وكما يلي:

```
>> normplot(x)
```



شكل ٥.٤ رسم البيانات ذات الصفوف العشوائية مِنْ التوزيعِ الأسّي بأستخدام الدالة normplot

٥.٣.٢ أختبار توزيع عينتين بأستخدام دالة الرسم quantile-quantile plot

دالة الرسم quantile-quantile تكون ذات فائدة فيما أذا كانت العينتين مأخوذتان مِنْ نفس التوزيعِ (سواء مُوَزَّعة طبيعيا أَو لا) . ففي المثال التالي تم أستخدام دالة poissrnd لتكوين عينتين من نفس البيئة وبقيم مختلفة ففي هذا المثال بالرغم من أنَّ حجوم العيّنةَ والمعالم مختلفة، فأن العلاقةَ ذات الخَطِّ المستقيم توضح بأنّ العينتين مؤخوذتان مِنْ نفس التوزيع. الشكل 5.5 أن مخطط الإحتمالِ الطبيعي ، الذي رسم quantile quantile لَه ثلاثة

عناصرِ تخطيطيةِ يمكن تمييزها بسهولة وهي الجزء الممثل بالرسم بالاشارة زائد (+) وهو يمثل quantiles لكُلّ عيّنة وبالأساس فأن عدد الإشارات هي عددُ قِيَمِ البياناتِ في العيّنةِ الأصغرِ. أما الجزء الممثل بالخَطُّ المستمر فهو يمثل الربط بين إلنسبة الخامسة والعشرين والنسبة الخامسة والسبعوي للعيناتِ. أما اجزء الثالث الممثل يالخط المتقطع يمثل أمتداد الخَطُّ المستمر إلى حدّ نهاية العيّنة. وبعدها نستخدم الدالة qqplot لرسم العينتين حيث تستخدم هذه الدالة لرسم أكثر من عينة على نفس المخطط وكما موضح في الشكل ٥.٥.

لتوليد العينة نفرض أن معدل القيم هو μ وأن حجم العينة هو N وأنها موجودة في عمود واحد هو C وعليه يمكن كتابة الدالة كما يلي:

x = poissrnd(μ, N,C)

وباختيار العينة الاولى بالمعالم μ=١٠ و N=١٥ و C=١ يمكن أن نكتب الدالة poissrnd على الشاشة الرئيسية كما يلي :

```
>> x = poissrnd(10,15,1)

x =

    18      13      14      9      16
     6      10       6      4       9
    10       8       5     13      10
```

وباختيار العينة الاولى بالمعالم μ=٥ و N=٢٥ و C=١ يمكن أن نكتب الدالة poissrnd على الشاشة الرئيسية كما يلي :

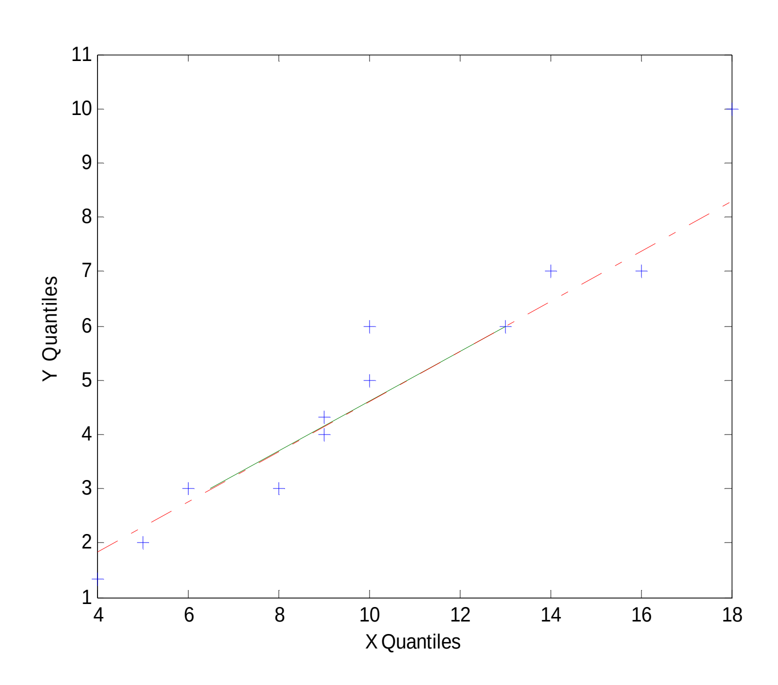
```
>> y = poissrnd(5,25,1)

y =

4       4       3       5       5
2       5       6       11      3
7       7       7       6       6
4       1       2       3       6
5       3       8       3       6
```

ثم بعد ذلك نستخدم الدالة qqplot لرسم العينتين x ، y وكما يلي:

```
>> qqplot(x,y);
```



شكل ٥.٥ رسم بياني لعينتين مأخوذتين مِنْ نفس التوزيعِ

أما أذا كانت العينتين مأخوذتان مِنْ مجتمعين مختلفتين في التوزيعِ ولتوضيح هذه الحالة نأخذ المثال التالي ونولد العينة الاولى x من مجتمع الصفوف العشوائية مِنْ التوزيعِ الطبيعيِ لخمس وعشرين قيمة وباختيار المعالم μ=٥ وσ=١ و N=٢٥ و C=١ حيث يمكن أن نكتب الدالة normrnd على الشاشة الرئيسية كما يلي :

```
>> x = normrnd(5,1,25,1)
```

x =				
5.6232	5.7990	5.9409	4.0079	5.2120
5.2379	3.9922	4.2580	6.0823	4.8685
5.3899	5.0880	4.3645	4.4404	5.4437
4.0501	5.7812	5.5690	4.1783	4.7344
3.8122	2.7977	5.9863	4.4814	5.3274

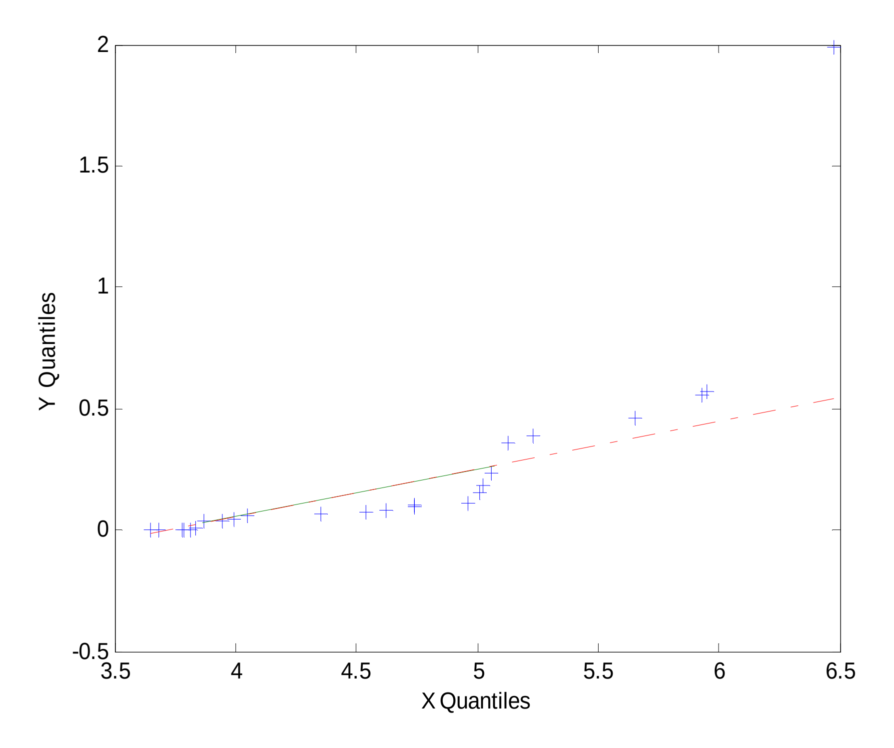
ونولد العينة الثانية y من بيئة الصفوف العشوائية مِنْ توزيعِ Weibull وكذلك لخمس وعشرون قيمة وباختيار المعالم μ=٢ وσ=٠.٥ و N=٢٥ و C=١ حيث يمكن أن نكتب الدالة weibrnd على الشاشة الرئيسية كما يلي :

```
>> y = weibrnd(2,0.5,25,1)
```

y =				
1.9913	0.0037	0.4642	0.3884	0.5598
0.0691	0.0818	0.0052	0.5744	0.3599
0.0709	0.0027	0.1046	0.0405	0.0381
0.0006	0.0556	0.1525	0.0471	0.0001
0.1866	0.0002	0.1078	0.0933	0.2352

ثم يمكن رسم العينتين معا على نفس المخطط كما موضحة في الشكل ٥.٦ باستخدام دالة qqplot وكما يلي:

```
>> qqplot(x,y)
```



شكل ٥.٦ رسم بياني لعينتين مأخوذتين مِنْ بيئتين مختلفتين

٥.٣.٣ أختبار توزيع عينتين من مجتمعين بأستخدام دالة الرسم

Weibull Probability Plots

دالة الرسم Weibull تكون ذات فائدة فيما أذا كانت العينتين مأخوذتان مِنْ توزيعِين مختلفين (سواء مُوَزَّعة طبيعيا أَو لا). العديد مِنْ تحليلات توضح فرضيةَ التوزيعَ لدالة Weibull ، لذا فأن هذا المخطط يُمْكِنُ أَنْ يزوّدَ بَعْض التأمينِ والثقة وِأَنَّ هذه الفرضيةِ لَمْ يتم أختراقها أو فقدان الثقة بها . يمكن أن تكون التقسيمات في مِقياس الإحداثي الصادي لَيست موحدة وإنّ قِيَمَ الإحداثي الصادي (y) تمثل الإحتمالاتَ، وهي تتراوح بين صفر والواحد وأن المسافة بين التقسيمات الزمنية على الإحداثي الصادي تطابق مع المسافات بين quantiles في توزيع Weibull . إذا كانت نقاطِ البياناتَ للأشارات زائد (+) تقع قُرْب الخَطِّ،

فأن الفرضية التي تفرض أن البياناتَ تتبع توزيع Weibull تكون قريبة جدا من هذا الافتراض.

المثال التالي وكما موضح في الشكل ٥.٧ يبين الصفوف العشوائية مِنْ توزيعِ Weibull بأستخدام الدالة weibrnd حيث تم أختيار العينة لخمس وعشرون قيمة ولعمود واحد ثم بعد ذلك تم رسم البيانات بأستخدام الدالة weibplot.

لتوليد العينة نفرض أن معالم العينة هما A و B وأن حجم العينة هو N وأنها موجودة في عمود واحد هو C وعليه يمكن كتابة الدالة كما يلي:

```
x = weibrnd (A,B ,N,C)
```

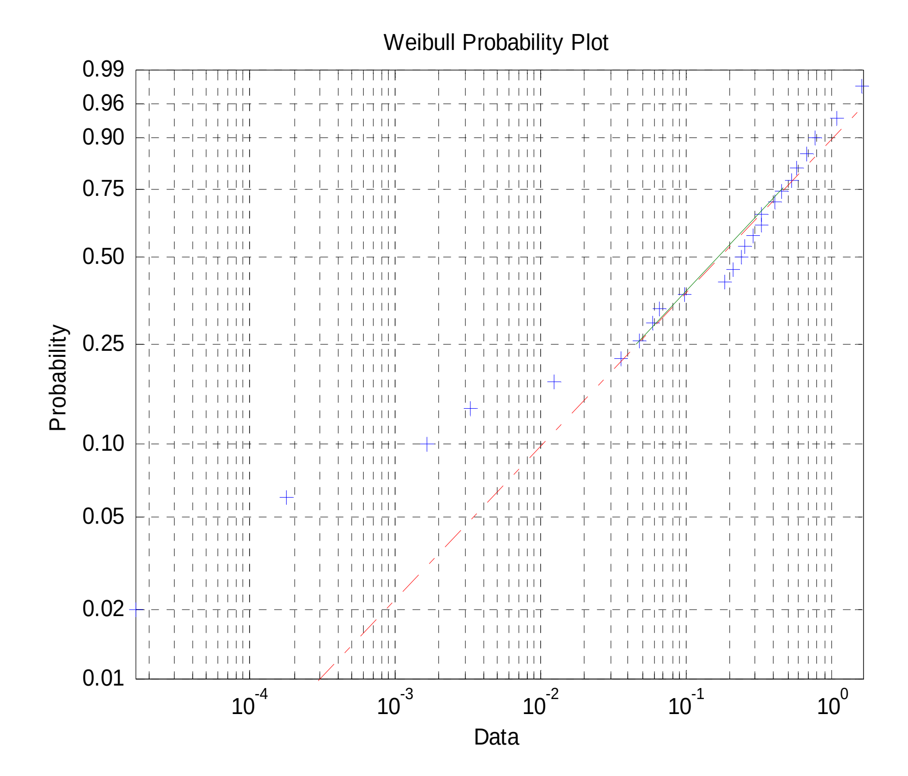
وباختيار A=2 وB=0.5 و N=٢٥ و C=١ يمكن أن نكتب الدالة weibrnd على الشاشة الرئيسية كما يلي :

```
>> x = weibrnd(2,0.5,25,1)
```

```
x =
0.0124    0.1850    0.0032    0.0000    0.2089
0.2899    0.3304    0.2549    0.0588    0.0663
0.0357    0.0002    0.4044    0.7646    0.0017
0.4530    0.2404    0.0482    0.3309    0.6791
0.5300    0.5737    1.6037    1.0904    0.0963
```

ثم بعد ذلك نرسم العينة x كما في الشكل ٧.٧ وبأستخدم الدالة weibplot وكما يلي:

```
>> weibplot(x)
```



شكل ٥.٧ رسم للبيانات العشوائية مِنْ توزيعِ Weibull بأستخدام الدالة weibplot

٥.٤ التوزيعِ التجميعي

Empirical Cumulative Distribution Function (CDF)

إذا لم تكن راغبا لإفتِراض بأنّ بياناتِكَ تتبع توزيع إحتمالِي معيّنِ ، يُمْكِنك أَنْ تَستعملَ الدالة cdfplot لرسم تخمينَ تجريبيَ مِنْ دالة التوزيعِ التجميعي (cdf) . تَحْسب هذه الدالة نسب نقاط البياناتِ التي تُشيرُ الى أقل مِنْ كُلّ قيمة x ، وترسم النسبةَ كدالة الى x وإنّ مِقياسَ الإحداثي الصادي يكون خطي . يوضح المثالِ التالي وظيفةَ التوزيعِ التجميعي التجريبي لعيّنة Weibull وكيفية رسمها بأستخدام الدالة cdfplot التي تستخدم لعرض

البيانات وكما موضحة في الشكل ٥.٨ حيث أن الرسم يوضح أن دالة الاحتمالِ تَرتفعُ بشكل حادّ قُرْب x =0 وتكون مستوية للقِيَمِ الأكبرِ. أكثر من ٨٠ % من القيم تكون أقل مِنْ ١ وأن القِيَمِ الباقيةِ نَشرتْ على المدى [٥ ١].

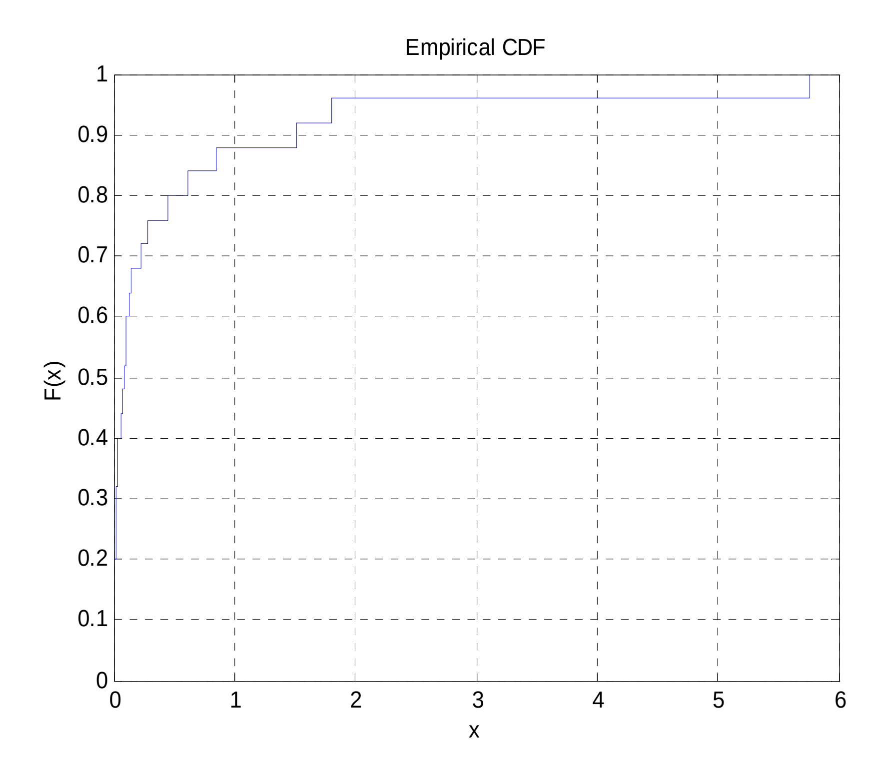
لتوليد العينة نختار معالمها وكما في المثال السابق A=2 وB=0.5 و N=٢٥ و C=١ يمكن أن نكتب الدالة weibrnd على الشاشة الرئيسية كما يلي :

```
>> x = weibrnd(2,0.5,25,1)
```

```
x =
0.0218    0.1279    0.0465    2.5216    0.2993
0.1748    1.3419    0.0369    3.2533    0.3231
2.1717    0.2477    0.2292    0.2976    0.0034
0.5147    1.5798    0.3594    0.7350    0.0279
0.0139    0.3051    0.0152    0.0160    1.2732
```

ثم بعد ذلك نرسم العينة x كما في الشكل ٥.٨ وبأستخدم الدالة cdfplot وكما يلي:

```
>> cdfplot(x)
```



شكل ٥.٨ رسم للبيانات العشوائية مِنْ توزيعِ Weibull بأستخدام الدالة cdfplot

٥.٥ رسومات التشتت Scatter Plots

مخطط التشتت يمثل رسم بسيط مِنْ متغيّرِ واحدة لأخروبأستخدام حزمة MATLAB وظائف التشتت يُمْكِنُ أَنْ تولد رسومات التشتت. أن دالة plotmatrix يُمْكِنُ أَنْ تُنتجَ مصفوفة لهذه الرسومات توضح العلاقةَ بين عِدّة أزواج من المتغيّراتِ. يُضيفُ صندوقُ العُدّة الإحصائية لحزمة MATLAB الوظائفَ التي تُنتجُ النسخَ المُجَمَّعةَ لهذه الرسومات حيث أن هذه الرسومات مفيدة لحساب فيما أذا كانت قِيَمِ المتغيرات أَو العلاقةِ بين تلك المتغيرات نفسها في كُلّ مجموعة.

فالمثال التالي يوضح فيما لو أردت أختبار الوزنِ والمسافة بالأميالِ للسياراتِ مِنْ ثلاث سَنَواتِ لنماذج مختلفةِ فأننا أولا نحمل الملف المعني بذلك وهو carsmall بأستخدام الدالة load وهو يمثل عينة حجمها مئة قيمة

وبعدها نرسم البيانات بأستخدام الدالة gscatter والتي تمثل مخطط التشتت بتجميع المتغيّراتِ كما موضحة في الشكل ٥.٩ وهذا يوضح أنه لَيسَ فقط هناك علاقة قوية بين وزنِ السيارة والمسافة بالأميال، لكن أيضاً تلك السياراتِ الأحدثِ تَمِيلُ إلى أن تكون أخف ولَها القابلية على قطع المسافة بالأميالُ أفضلِ مِنْ السياراتِ الأقدمِ.

```
>> load carsmall
>> gscatter(Weight,MPG,Model_Year,'','xos')
```

كتوضيح سوف نعرض المتغيرات المستخدمة في هذه العينة والمخزونة في الملف ومن هذه المتغيرات أولا الوزن (Weight) وهي مئة قيمة

```
>> Weight

     Weight =

      3504    3693    3436    3433    3449    4341    4354
      4312    4425    3850    3090    4142    4034    4166
      3850    3563    3609    3353    3761    3086    2372
      2833    2774    2587    2130    1835    2672    2430
      2375    2234    2648    4615    4376    4382    4732
      2464    2220    2572    2255    2202    4215    4190
      3962    4215    3233    3353    3012    3085    2035
      2164    1937    1795    3651    3574    3645    3193
      1825    1990    2155    2565    3150    3940    3270
      2930    3820    4380    4055    3870    3755    2605
      2640    2395    2575    2525    2735    2865    3035
      1980    2025    1970    2125    2125    2160    2205
      2245    1965    1965    1995    2945    3015    2585
      2835    2665    2370    2950    2790    2130    2295
      2625    2720
```

عدد الاميال التي تقطعها السيارة في الجالون (MPG)

```
>> MPG

  MPG =

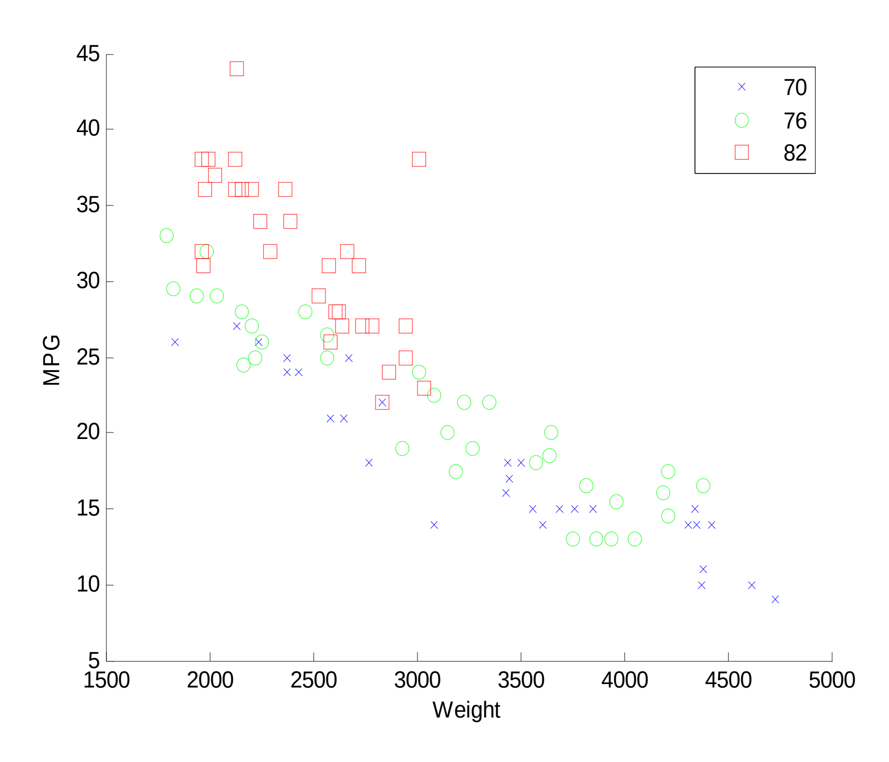
  18.0000   15.0000   18.0000   16.0000   17.0000   15.0000   14.0000
  14.0000   14.0000   15.0000       NaN       NaN       NaN       NaN
      NaN   15.0000   14.0000       NaN   15.0000   14.0000   24.0000   22.0000
  18.0000   21.0000   27.0000   26.0000   25.0000   24.0000   25.0000
  26.0000   21.0000   10.0000   10.0000   11.0000    9.0000   28.0000   25.0000
  25.0000   26.0000   27.0000   17.5000   16.0000   15.5000   14.5000
  22.0000   22.0000   24.0000   22.5000   29.0000   24.5000   29.0000
  33.0000   20.0000   18.0000   18.5000   17.5000   29.5000   32.0000
  28.0000   26.5000   20.0000   13.0000   19.0000   19.0000   16.5000
  16.5000   13.0000   13.0000   13.0000   28.0000   27.0000   34.0000
  31.0000   29.0000   27.0000   24.0000   23.0000   36.0000   37.0000
  31.0000   38.0000   36.0000   36.0000   36.0000   34.0000   38.0000
  32.0000   38.0000   25.0000   38.0000   26.0000   22.0000   32.0000
  36.0000   27.0000   27.0000   44.0000   32.0000   28.0000   31.0000
```

سنوات صنع السيارة أو موديل السيارة (Model_Year)

```
>> Model_Year

  Model_Year =

   70   70   70   70   70   70   70   70   70   70   70   70   70   70   70
   70   70   70   70   70   70   70   70   70   70   70   70   70   70   70
   70   70   70   70   70   76   76   76   76   76   76   76   76   76   76
   76   76   76   76   76   76   76   76   76   76   76   76   76   76   76
   76   76   76   76   76   76   76   76   76   82   82   82   82   82   82
   82   82   82   82   82   82   82   82   82   82   82   82   82   82   82
   82   82   82   82   82   82   82   82   82   82
```



شكل ٥.٩ رسم البيانات لعينة مختلفة بأستخدام الدالة gscatter

كذلك تَحتوي مجموعةُ معلومات carsmall على متغيّرات أخرى والتي تضيف سماتَ مختلفةَ أخرى للسياراتِ. حيث يُمْكِنك أَنْ تَختبر عِدد منها بعرض واحد وذلك بتكوين مخطط مصفوفةَ المجموعة. فالجهة العليا اليمنى للرسم الثانوي يوضح (MPG) مقابل القوة الحصانية، ويعرض ذلك على مدى السنين، حيث أن القوة الحصانية للسياراتِ قلت لكن المسافة التي تقطعها بالأميالَ تَحسّنَت. والدالة gplotmatrix تستحدم لمصفوفة الرسم المشتت بالمتغيرات المجمعة حيث يُمْكِنُ أَنْ ترسم كُل الأزواج أيضاً مِنْ قائمة واحدة مِنْ المتغيّراتِ سويّة مع المدرج الإحصائي لكُل متغيّر. ثم بعد ذلك نستخدم الدالة gplotmatrix لرسم المتغيرات وكما موضحة في الشكل 5.10.

فبعد أن تم تحميل الملف carsmall يمكن سحب متغيرات أخرى منه الوزن (Weight) مع المسافة (Displacement) مع القوة الحصانية (Horsepower) ونضعها في عينة xvars وكما مرتبة في الاعمدة حيث العمود الأول يمثل الوزن والعمود الثاني يمثل المسافة والعمود الثالث يمثل القوة الحصا نية وكما يلي:

```
>> xvars = [Weight Displacement Horsepower]

xvars =

3504    307    130
3693    350    165
3436    318    150
3433    304    150
3449    302    140
4341    429    198
4354    454    220
4312    440    215
4425    455    225
3850    390    190
3090    133    115
4142    350    165
4034    351    153
4166    383    175
3850    360    175
3563    383    170
3609    340    160
3353    302    140
3761    400    150
3086    455    225
```

2372	113	95
2833	198	95
2774	199	97
2587	200	85
2130	97	88
1835	97	46
2672	110	87
2430	107	90
2375	104	95
2234	121	113
2648	199	90
4615	360	215
4376	307	200
4382	318	210
4732	304	193
2464	107	86
2220	116	81
2572	140	92
2255	98	79
2202	101	83
4215	305	140
4190	318	150
3962	304	120
4215	351	152
3233	225	100
3353	250	105
3012	200	81
3085	232	90

2035	85	52
2164	98	60
1937	90	70
1795	91	53
3651	225	100
3574	250	78
3645	250	110
3193	258	95
1825	97	71
1990	85	70
2155	97	75
2565	140	72
3150	130	102
3940	318	150
3270	120	88
2930	156	108
3820	168	120
4380	350	180
4055	350	145
3870	302	130
3755	318	150
2605	112	88
2640	112	88
2395	112	88
2575	112	85
2525	135	84
2735	151	90
2865	140	92

3035	151	NaN
1980	105	74
2025	91	68
1970	91	68
2125	105	63
2125	98	70
2160	120	88
2205	107	75
2245	108	70
1965	91	67
1965	91	67
1995	91	67
2945	181	110
3015	262	85
2585	156	92
2835	232	112
2665	144	96
2370	135	84
2950	151	90
2790	140	86
2130	97	52
2295	135	84
2625	120	79
2720	119	82

فبعد أن تم تحميل الملف carsmall يمكن سحب متغيرات أخرى منه مثل عدد الاميال المقطوعة لكل جالون مع التعجيل ونضعها في عينة yvars وكما مرتبة في الاعمدة حيث العمود الاول يمثل الاميال المقطوعة والعمود الثاني يمثل التعجيل وكما يلي:

```
>> yvars = [MPG Acceleration]

   yvars =

   18.0000   12.0000
   15.0000   11.5000
   18.0000   11.0000
   16.0000   12.0000
   17.0000   10.5000
   15.0000   10.0000
   14.0000    9.0000
   14.0000    8.5000
   14.0000   10.0000
   15.0000    8.5000
       NaN   17.5000
       NaN   11.5000
       NaN   11.0000
       NaN   10.5000
       NaN   11.0000
   15.0000   10.0000
   14.0000    8.0000
       NaN    8.0000
   15.0000    9.5000
   14.0000   10.0000
   24.0000   15.0000
```

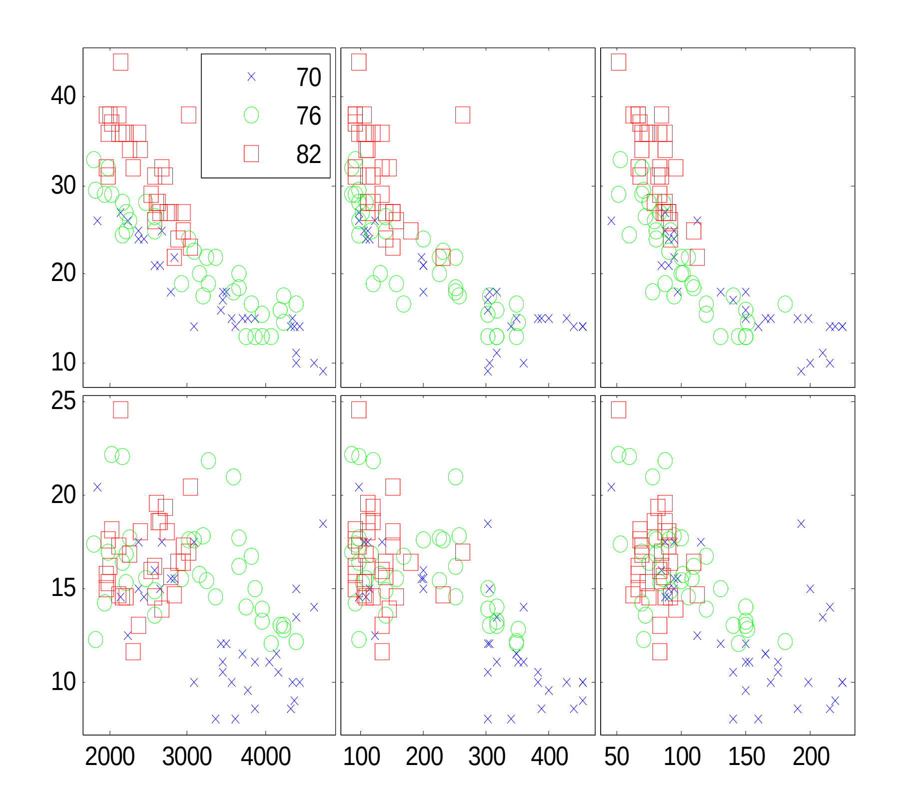
22.0000	15.5000
18.0000	15.5000
21.0000	16.0000
27.0000	14.5000
26.0000	20.5000
25.0000	17.5000
24.0000	14.5000
25.0000	17.5000
26.0000	12.5000
21.0000	15.0000
10.0000	14.0000
10.0000	15.0000
11.0000	13.5000
9.0000	18.5000
28.0000	15.5000
25.0000	16.9000
25.0000	14.9000
26.0000	17.7000
27.0000	15.3000
17.5000	13.0000
16.0000	13.0000
15.5000	13.9000
14.5000	12.8000
22.0000	15.4000
22.0000	14.5000
24.0000	17.6000
22.5000	17.6000
29.0000	22.2000

24.5000	22.1000
29.0000	14.2000
33.0000	17.4000
20.0000	17.7000
18.0000	21.0000
18.5000	16.2000
17.5000	17.8000
29.5000	12.2000
32.0000	17.0000
28.0000	16.4000
26.5000	13.6000
20.0000	15.7000
13.0000	13.2000
19.0000	21.9000
19.0000	15.5000
16.5000	16.7000
16.5000	12.1000
13.0000	12.0000
13.0000	15.0000
13.0000	14.0000
28.0000	19.6000
27.0000	18.6000
34.0000	18.0000
31.0000	16.2000
29.0000	16.0000
27.0000	18.0000
24.0000	16.4000
23.0000	20.5000

```
36.0000   15.3000
37.0000   18.2000
31.0000   17.6000
38.0000   14.7000
36.0000   17.3000
36.0000   14.5000
36.0000   14.5000
34.0000   16.9000
38.0000   15.0000
32.0000   15.7000
38.0000   16.2000
25.0000   16.4000
38.0000   17.0000
26.0000   14.5000
22.0000   14.7000
32.0000   13.9000
36.0000   13.0000
27.0000   17.3000
27.0000   15.6000
44.0000   24.6000
32.0000   11.6000
28.0000   18.6000
31.0000   19.4000
```

ثم بعد ذلك نرسم العينة كما موضحة في الشكل ٥.١٠ بأنستخدم الدالة gplotmatrix وكما يلي:

```
>> gplotmatrix(xvars,yvars,Model_Year,'','xos')
```



شكل ٥.١٠ رسم البيانات للسماتَ المختلفةَ للسياراتِ بأستخدام الدالة gplotmatrix

الفصل السادس

تطبيقات حزمة MATLAB

في اختبارات الفرضية

Application of MATLAB

in Hypothesis Testes

٦.١ مقدمة	Introduction
٦.٢ مصطلح اختبار الفرضية	Hypothesis Test Terminology
٦.٣ اختبارات الفرضية	Hypothesis Tests
٦.٤ دوال أختبار الفرضيات	Hypothesis Test Functions

- ٦.٤.١ أختبار jb test
- ٦.٤.٢ أختبار kstest
- 6.4.3 أختبار kstest2
- ٦.٤.٤ أختبار lillietest
- 6.4.5 أختبار ranksum
- ٦.٤.٦ أختبار signrank
- ٦.٤.٧ أختبار ttest
- ٦.٤.٨ أختبار ttest٢
- ٦.٤.٩ أختبار ztest

٦.١ مقدمة Introduction

توجد طرق مختلفة ومتنوعة لاختبارات الفرضية على عينة معينة أو مجموعة من العينات من نفس البيئة أو من بيئات مختلفة لحساب مدى تقدير الفرضية وصحتها أو قربها من النتيجة الحقيقية. لهذه الطرق أهمية كبيرة في نظم دعم القرارات وكيفية التوصل الى أختيار القرار الصحيح من بين مجموعة القرارات والاحتمالات الموجودة وذلك بأتباع الخطوات العلمية الصحيحة لكي نتوصل الى النتيجة المطابقة أعتمادا على البيانات المتوفرة.

سوف نستعرض في هذا الفصل وصفا مختصرا لاختبارات الفرضية وكيفية تطبيقها باستخدام حزمة MATLAB حيث يوجد حقل كامل ضمن التطبيقات الاحصائية في حزمة MATLAB يعنى بهذا الموضوع وتطبيقاته. تعرف اختبارات الفرضية بأنها مجموعة من الاختبارات والإجراءات تستخدم لحساب وتقدير الافتراض لخواص عينة معينة وهل يكون هذا الافتراض معقول أو صحيح أم لا. وتوضيحا لذلك نستعرض المثال التالي، يفترض أن شخصا ما يقول أن سعر لتر البنزين الخالي من الرصاص في ولاية كاليفورنيا هو ١.٢ دولار فكيف تقرر أن هذا البيان حقيقي؟ للتحقيق من ذلك فأن يجب مراقبة محطات التعبئة في كاليفورنيا وكم يباع لتر البنزين في كل محطة وما هو نوع البنزين وبعد جمع المعلومات بالإمكان التحقق من كون هذا البيان صحيح. وكذلك بالإمكان اخذ عدد من المحطات بشكل عشوائي وحساب معدل سعر اللتر ومقارنته بالسعر ١.٢ دولار وأن هذه النتيجة من المحتمل لا تساوي بالضبط ١.٢ دولار بل مقاربة لها فأنها كذلك تعتمد في العمل.

مثال أخر لو أن شخصا ما يقول أن سعر كيلو السكر في عمان هو ٤٠ قرشا فكيف نتحقق من ذلك؟ عملية التحقق من ذلك بمراقبة كافة المحلات والمؤسسات التجاريةالتي تبيع السكر عملية صعبة جدا ولا يمكن القيام بها أو تطبيقها بالشكل الدقيق والصحيح لان ذلك يتطلب منا جهدا كبيرا ويتطلق الاستعانة بكادر كبيرأضافة الى أنفاق مبالغ مالية كبيرة لتهيأة البيانات ومتابعتها. لذا نلجأ إلى الحل الثاني وهو أخذ عينة عشوائية من المحلات والمؤسسات ونحسب معدل سعر الكيلو من السكرومقارنته بـ ٤٠ قرشا حيث من

المحتمل أن تكون النتيجة ليست مساوية بالضبط ٤٠ قرشا وإنما مقاربـة لهـا وبـذلك تكون قد تحققت فرضيتنا.

٦.٢ مصطلح اختبار الفرضية Hypothesis Test Terminology

هنـاك مجموعـة مـن المصـطلحات التـي تتعلـق بأختبـار الفرضـية وسـوف نستعرضها وكما يلي:

١. فرضية العدم (Null Hypothesis): فرضية العدم تعني التقدير الأصلي أي أن معدل سعر لتر البنزين في المثال الأول يكون مساوي بالضبط إلى ١.٢ دولار وأن الدلالة تكتب كما يلي:

$$Ho: \mu = 1.2$$

٢. الفرضية البديلة (Alternative Hypothesis): وتوجد ثلاث أنواع من الفرضية البديلة وهي:

- الاحتمال الأول إذا كنت مهتم بالأسعار فقـط وكـان سـعر لـتر البنـزين في المثال الأول في الواقع أعلى في هذه الحالة فأن الفرضية البديلة تكتب:

$$H1: \mu > 1.2$$

- الاحتمال الثاني إذا كان سعر لتر البنزين في المثال الأول أقل فـأن الفرضـية البديلة تكتب:

$$H1: \mu < 1.2$$

- الاحتمال الثالث: إذا كان سعر لتر البنزين في المثال الأول لا يساوي القيمة المفترضة لذلك فأن الفرضية تكتب:

$$H1: \mu \neq 1.2$$

٣. المستوى الفعال (Significance Level): المستوى الفعـال أو ذو الأهميـة يعنـي درجة التأكد المطلوبة لكي ترفض فرضية العدم لمصلحة البديل. بأخذ عينـة صـغيرة لا يمكنك أن تتأكد من استنتاجاتك لذا فإن عليك أن تقرر لاحقاً لرفض فرضية العدم إذا كانت احتمالية نتائجـك المتحققـة أقـل مـن المسـتوى الفعـال أو المسـتوى المطلـوب. للأمور القياسية تستخدم قيمة ٥% ، فإذا كانت قيمة النتائج المستخلصة هي بمقـدار ٥% من القيمة المفترضة فأنه لا يمكن رفض القيمـة، أمـا إذا أردت اكـثر دقـة فتختـار قيمة α أقل من ٥% لتكون النتائج أدق وهكذا.

٤. قيمة P (P Value): هي احتمالية مراقبة نتيجـة العينـة المعطـاة تحـت فرضـية معينة على أن فرضية العدم هي صحيحة. إذا كانت قيمة P أقل مـن α فعنـد ذلـك نرفض فرضـية العـدم وكمثـال عـلى ذلـك إذا كانـت α = ٠.٠٥ و P = ٠.٠٣ فعليـه سوف ترفض فرضية العدم، والحدث يكون غير صحيح. أمـا إذا كانـت P أكـبر مـن α فأنه لديك دليل غير كافٍ لرفض فرضية العدم.

٥. فترات الثقة (Confidence Intervals): تستخدم فترات الثقـة لاختبـار العديـد من النتائج، وأن فترات الثقة تمثل مدى القيم التي لـديها احـتمالات مختـارة لاحتـواء كمية من الفرضيات. كمثال على ذلك لو فرضنا أن قيمة ٠.٥ في داخل ٥% من فـترات الثقة للمعدل M هذا مشابه للمستوى الفعال أو المهم ٠.٠٥ وبعكس ذلك أن (α – ١) ١٠٠ من فترات الثقـة لا تحتـوي عـلى نصـفه إذا سـوف تـرفض فرضـية العـدم في مستوى المحدد.

6.3 اختبارات الفرضية hypothesis Tests

أن الاختلاف بين إجراءات اختبار الفرضية في أغلب الأحيان ينشأ عن الاختلافات في الفرضيات التي يضعها الباحث حول عينة البيانات. فمثلاً أن اختبار Z يفرض أن البيانات تمثل عينات مستقلة من نفس التوزيع الطبيعي حيث وبالأمكان معرفة الانحراف المعياري لها. وأن اختبار T له نفس الفرضيات فيما عدا تخمين الانحراف المعياري الذي يستخدم بدلاً من تحديده ككمية معروفة.

الاختبارين أعلاه لهما نسبة من الخطأ يقـاس بنسـبة الإشـارة إلى الضوضـاء (S/N) وهي:

$$(S/N)_Z = \frac{\bar{X} - M}{\sigma}$$

$$(S/N)_T = \frac{\bar{X} - M}{S}$$

حيث أن

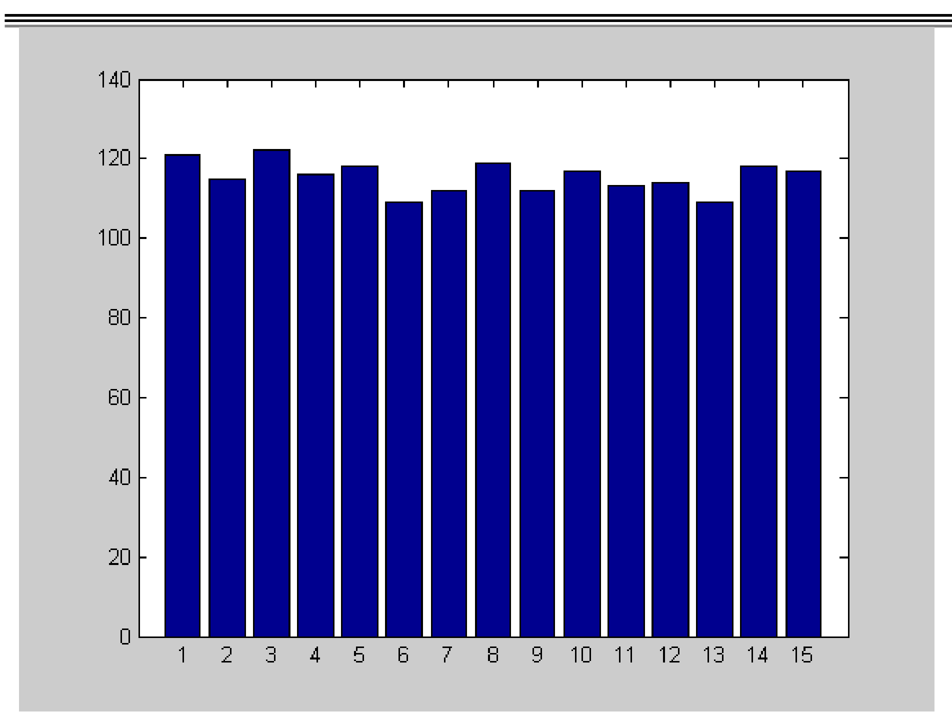
$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

وكمثال على ذلك نأخذ قيمة أسعار الموز في مدينة عمان فلو تم تسجيل أسعار المـوز في المدينة ومراقبة ذلك وعلى مدى شـهرين، فلـو فرضـنا أن البيانـات المستحصـلة في الشهر الاول للأسعار كانت كما يلي:

```
>> price1=[121 115 122 116 118 109 112 119 112 117 113
           114 109 118 117]
```

يمكن رسم هذه البيانات وذلك باسـتخدام الدالـة bar وكـما موضـحة في الشـكل ٦.١ وتكتب الدالة كما يلي:

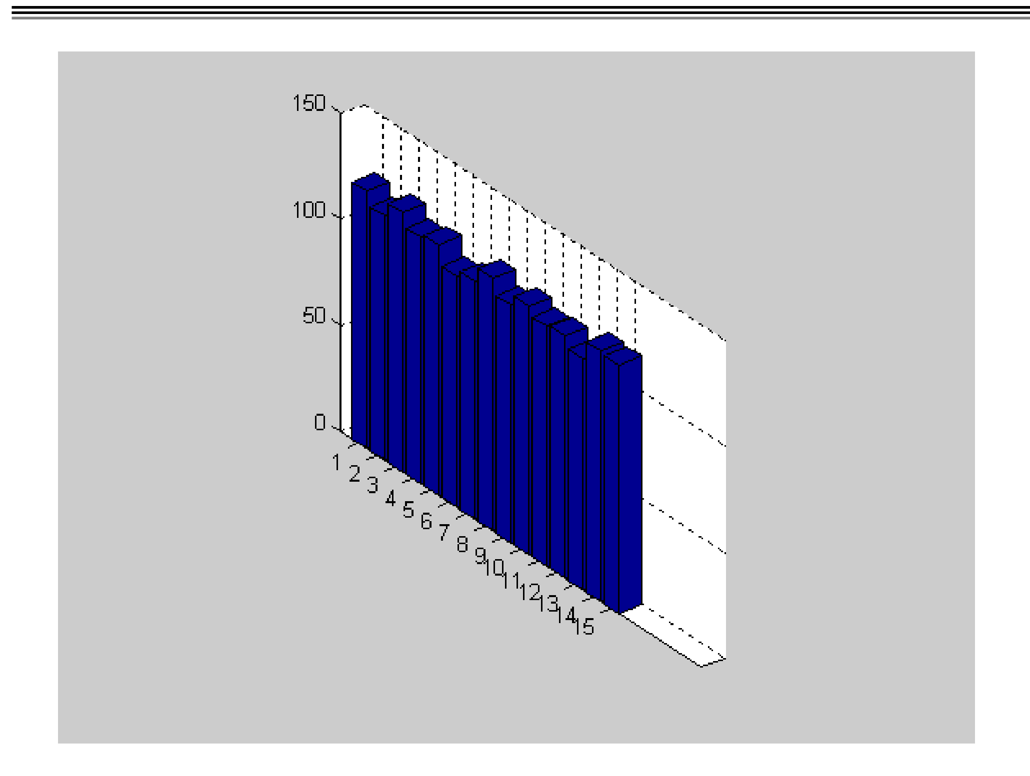
```
>> bar(price1)
```



شكل ٦.١ توزيع أسعار الموز في مدينة عمان للشهر الاول بأستخدام الدالة bar

يمكن أعادة رسم البيانات في الشكل ٦.١ بأستخدام صيغة أخرى مشابهة للدالة bar وذالك بأستخدام الدالة bar3 والتي تعطـي الرسـم بلأبعـاد الثلاثـة وكـما موضـحة في الشكل ٦.٢ ويكتب أمر هذه الدالة كما يلي:

```
>> bar3(price1)
```



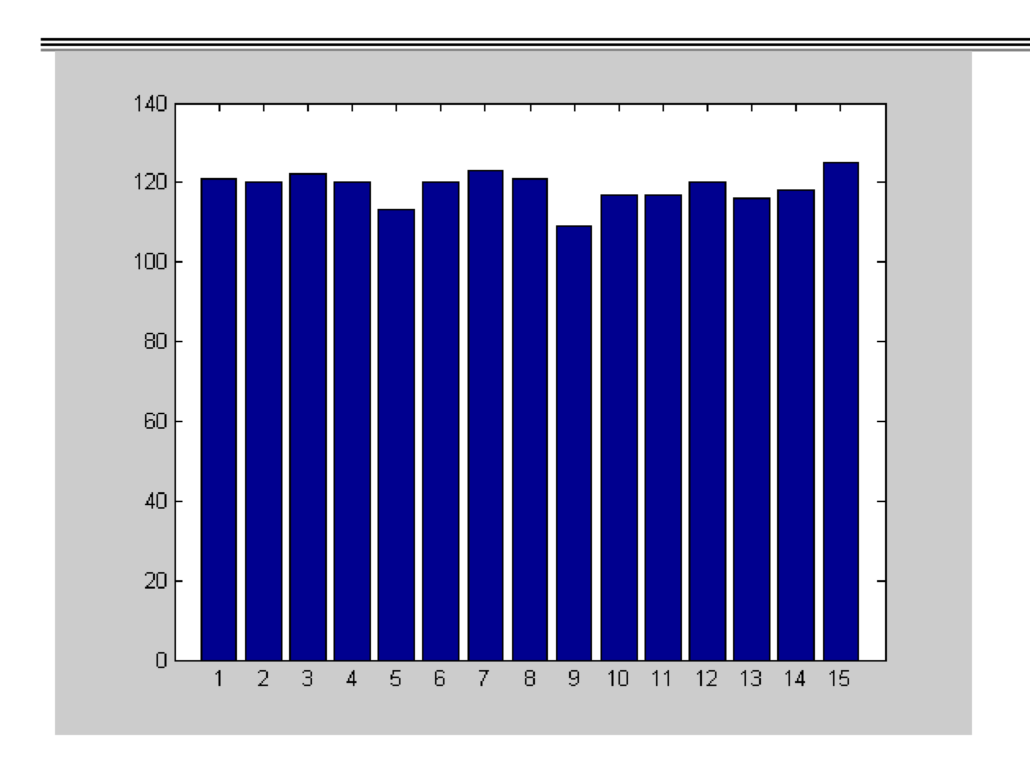
شكل ٦.٢ توزيع أسعار الموز في مدينة عمان للشهر الاول بأستخدام الدالة bar3

وفي الشهر الثاني أعدنا عملية المراقبة وتسجيل اسعار الموز لمدينة عمان حيث كانت قيمة ألأسعار المسجلة لهذا الشهر كما يلي:

```
>> price2=[121 120 122 120 113 120 123 121 109 117 117
          120 116 118 125]
```

يمكن رسم البيانات باستخدام الدالة bar وكما موضحة في الشكل ٦.٣ وتكتب الدالة كما يلي:

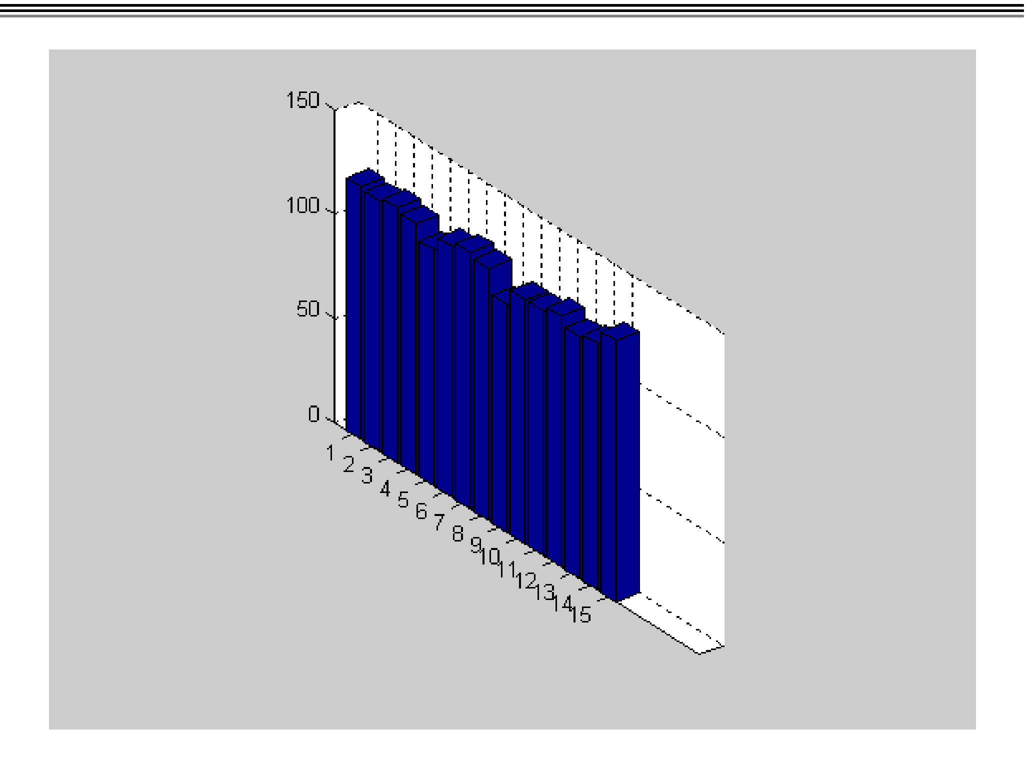
```
>> bar(price2)
```



شكل ٦.٣ توزيع أسعار الموز في مدينة عمان للشهر الثاني بأستخدام الدالة bar

يمكن أعادة رسم البيانات في الشكل ٦.٣ بأستخدام صيغة أخرى مشابهة للدالة bar وذالك بأستخدام الدالة bar3 والتي تعطـي الرسـم بلأبعـاد الثلاثـة وكـما موضـحة في الشكل ٦.٤ ويكتب أمر هذه الدالة كما يلي:

```
>> bar3(price2)
```



شكل ٦.٤ توزيع أسعار الموز في مدينة عمان للشهر الثاني بأستخدام الدالة bar3

كخطوة أساسية يجب معرفة فيما إذا كانت العينة لكل شهر تتبع التوزيع الطبيعي، وبما أن حجم العينة المستخدمة صغيرة يمكن أن نختار اختبار (lilliefors) وذلك باستخدام الدالة lillietest وكما يلي:

```
>> lillie Test  ( price 1 )

ans   =

          0

>> lillie Test  ( price 2 )

ans   =

          0
```

وبما أن النتيجة في الحالتين تساوي صفر فأنه ليس هناك داعي لرفض فرضية العدم وأن العينتين تكونان ذات توزيع طبيعي.

مثال على ذلك لو كانت قيمة P= 0.001، ذلك يعني أن احتمالية استحصال قيم كل من T, Z يكون واحد من ألف، وهذا يجعل الموضوع به شك حول فرضية العدم التي ترفض بدلا من اعتقادك أن النتيجة مطابقة ٩٩٩ وتبقى واحد.

مثال اخر نستعرض سعر لتر البنزين في الولايات المتحدة للشهرين الأول والثاني من عام ١٩٩٣ حيث أن هذه البيانات مخزونة في الملف gas.mat وهو موجود في التطبيقات الاحصائية لحزمة MATLAB ولتطبيق ذلك أولاً نحمل الملف gas بأستخدام الدالة load وكما يلي:

```
>> load gas
```

ثم بعد ذلك نعرض البيانات التي بداخل الملف وهي الأسعار للشهر الأول والأسعار للشهر الثاني ونفسر القيم بتحويل البيانات للشهرين الى مصفوفة من عمودين حيث يمثل العمود الاول أسعار البنزين للشهر الاول ويمثل العمود الثاني أسعار البنزين للشهر الثاني وكما يلي:

```
>> prices  = [ price 1     price 2]

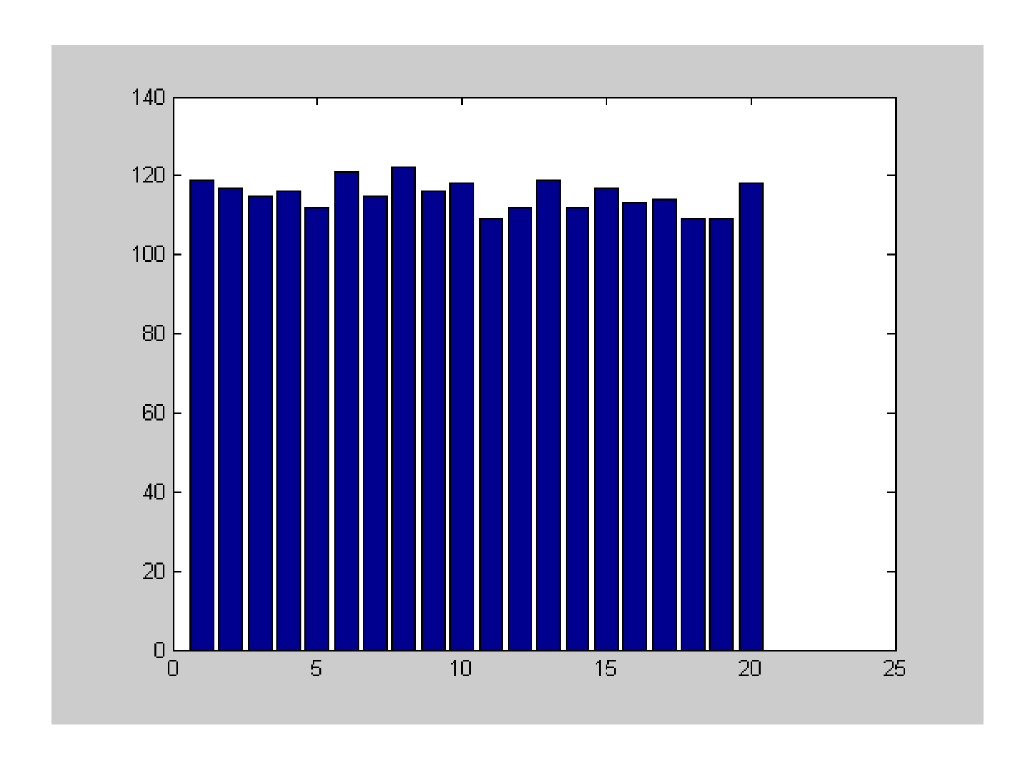

>> prices=[price1 price2]

prices =
1st month      2nd month
   119           118
   117           115
   115           115
```

116	122
112	118
121	121
115	120
122	122
116	120
118	113
109	120
112	123
119	121
112	109
117	117
113	117
114	120
109	116
109	118
118	125

فلو أردنا ملاحظة التوزيع للعينة الاولى والتي تمثل الاسعار للشهر الاول فأنه يمكن رسم العينة وذلك بأستخدام الدالة bar وكما موضحة في الشكل ٦.٥ وكما يلي:

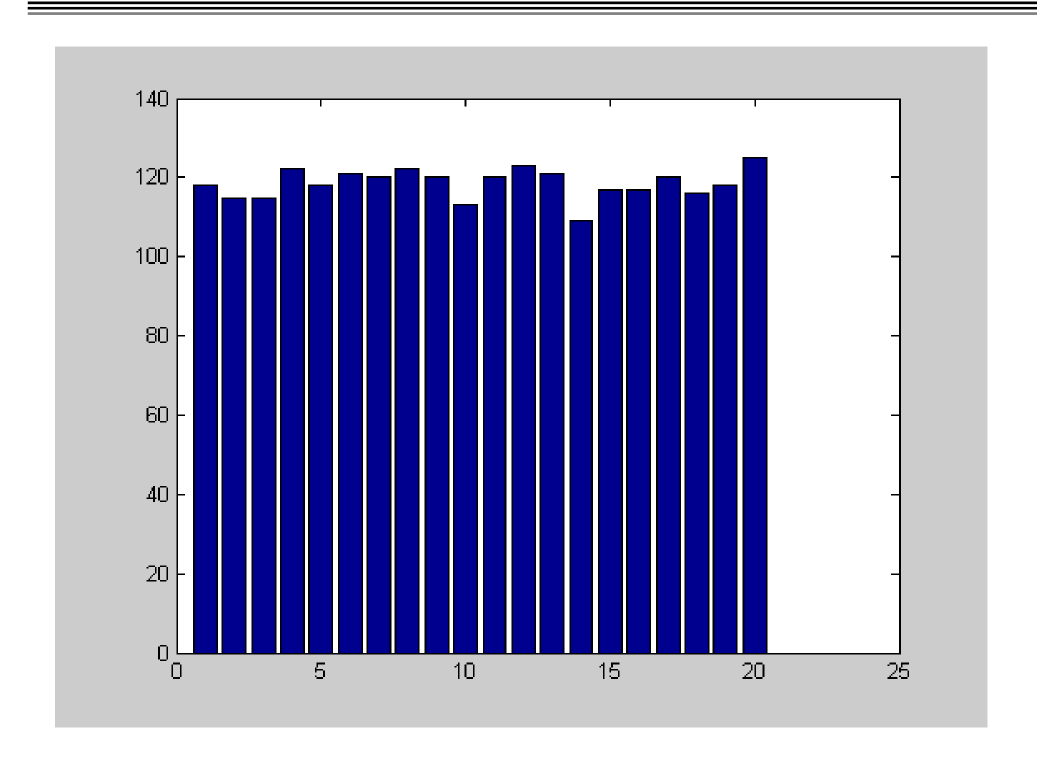
```
>> bar(price1)
```



شكل ٦.٥ توزيع عينة اسعار الموز للشهر الاول بأستخدام الدالة bar

وبنفس الطريقة يمكن رسم توزيع أسعار الموز للشهر الثاني وكما موضحة في الشـكل ٦.٦ وكما يلي:

```
>> bar(price2)
```



شكل ٦.٦ توزيع عينة اسعار الموز للشهر الثاني بأستخدام الدالة bar

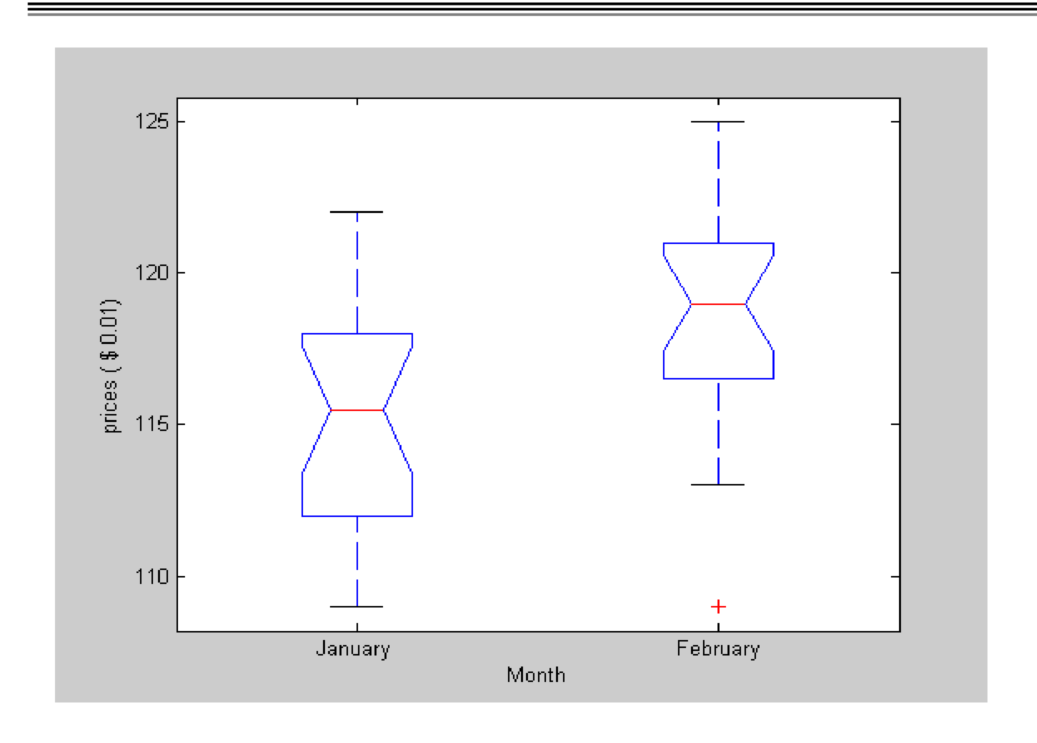
إن لم تكن العينتين موزعة طبيعياً ولكن لها نفس الشكل لذا يكون من الأفضل استخدام اختبار مجموعة الرتبة (ranksum) بدلاً من اختبار T . الدالة (ranksum) في حزمة MATLAB تقوم بأرجاع النتائج إلى اختبار الفرضية وتعتمد القيمة ٠.٠٥ كمستوى للاختبار حيث أن قيمة P هي نتيجة الاختبار وأن قيمة H لها حالتين، أما إن تساوي صفر وعندها تكون الأوساط متساوية وأما أن تساوي ١ وعندها تكون الأوساط غير متساوية. أما الحالة stats فأنها تحتوي على قيمة zval وهي عبارة عن القيمة الأحصائية الطبيعية لقيمة Z وهنا تساوي 2.5928- وأما قيمة ranksum فهي تساوي هنا 314 وأن استخدام الدالة ranksum يكون بكتابتها كما يلي:

```
>> [ P, h , stats ] = ranksum ( price1, price2 )
   P   = 0.0095
   h  = 1

                    Stats  =
          zval: -2.5928
          ranksum: 314
```

وباستخدام الدالة boxplot يمكن الوصول إلى نفس النتيجة ومعرفة التغيرات في الأسعار خلال الشهرين التي تم فيهما القياس وكما موضحة في الشكل ٦.٧ حيث نلاحظ بشكل عام أن مستوى الاسعار في الشهر الثاني أعلى من مستوى الاسعار في الشهر الاول. ويمكن تنفيذ الدالة boxplot وكذلك استخدام الدالة set لعرض الخواص ويتم ذلك كما يلي:

```
>> boxplot   ( prices, 1 )
>> set( gca,'xticklabel' , Str2mat('January' , 'February' ))
>> xlabel ('Month')
>> ylabel ('prices ( $ 0.01)')
```



شكل ٦.٧ التغيرات في أسعار البنزين للشهرين الاول والثاني بأستخدام الدالة boxplot

6.4 دوال أختبار الفرضيات Hypothesis Test Functions

توجد أنواع مختلفة من دوال أختبار الفرضيات في البرامج الاحصائية لحزمة MATLAB وسوف نتطرق في هذه الفقرة الى أنواعها وتطبيقاتها:

٦.٤.١ أختبار jbtest

وهو اختبار Jarque Bera ويستخدم لاختبار التوزيع الطبيعي لعينة واحدة. فإذا كانت قيمة h=1 فبإمكانك أن ترفض الفرضية بكون العينة ذات توزيع طبيعي أما إذا كانت h = 0 فلا يمكنك أن ترفض الفرضية أي أن العينة ذات توزيع طبيعي.

مثالاً على ذلك نحمل الملف carsmall وهو ملف مخزون ضمن التطبيقـات الاحصـائية ضمن حزمة MATLAB والذي يحتوي على معلومات مختلفة عـن السـيارات الصـغيرة ويعطي كذلك أوزانها ولمئة قيمة ويمكن تحميل الملف بأستخدام الدلة laod وكما يلي:

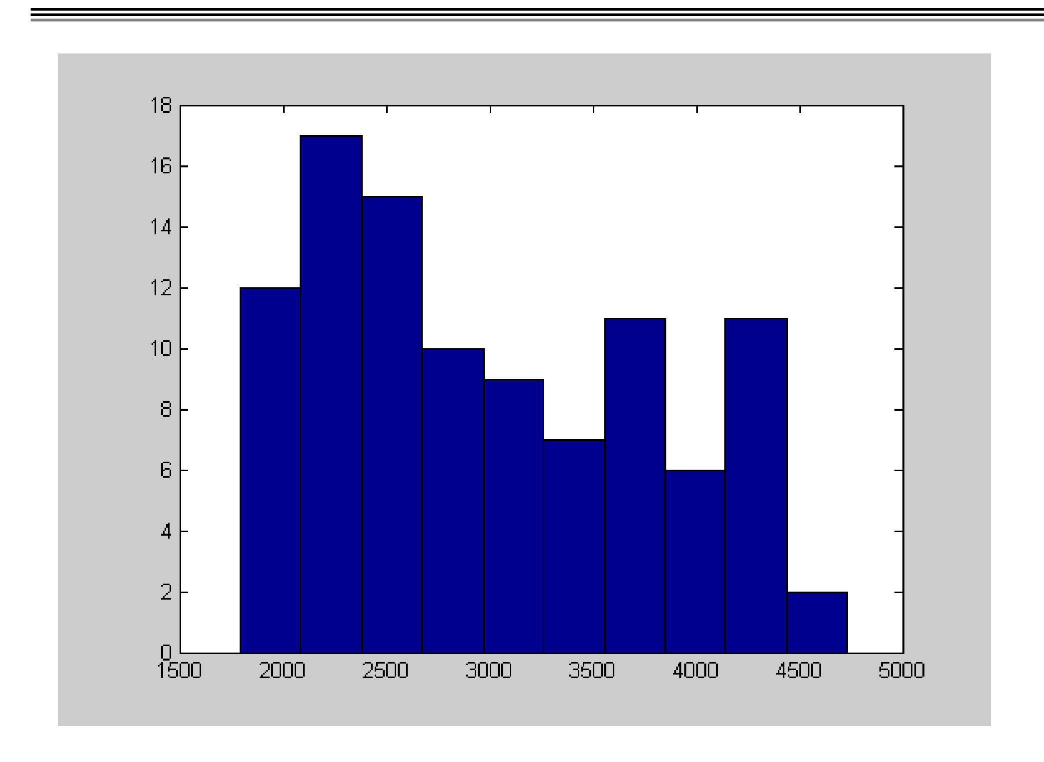
>> **load carsmall**

إذا أردنا أن نستعرض قيم الأوزان وهي مئة قيمة نكتب المتغير كما يلي:

```
>> Weight =
3504    3693    3436    3433    3449    4341
4354    4312    4425    3850    3090    4142
4034    4166    3850    3563    3609    3353
3761    3086    2372    2833    2774    2587
2130    1835    2672    2430    2375    2234
2648    4615    4376    4382    4732    2464
2220    2572    2255    2202    4215    4190
3962    4215    3233    3353    3012    3085
2035    2164    1937    1795    3651    3574
3645    3193    1825    1990    2155    2565
3150    3940    3270    2930    3820    4380
4055    3870    3755    2605    2640    2395
2575    2525    2735    2865    3035    1980
2025    1970    2125    2125    2160    2205
2245    1965    1965    1995    2945    3015
2585    2835    2665    2370    2950    2790
2130    2295    2625    2720
```

بالامكان أستخدام الدالة hist لايجاد توزيع أوزان السيارات والتكرار الذي يحصـل في الاوزان حسب العينة المختارة وكما موضحة في الشكل 6.8 ويكون الامر كما يلي:

>> hist(Weight)



شكل ٦.٨ التوزيع التكراري لعينة أوزان السيارات بأستخدام الدالة hist

والان نطبق اختبار jbtest على قيم الأوزان وكما يلي:

```
>> [ h , p , j ]  =  jbtest ( Weight)
    h  =  1
    p  = 0.026718
    j  =  7.2448
```

وبما أن قيمة h = 1 فأن الفرضية ترفض على كونها طبيعية وأن قيمة p = 2.67% كذلك نرفض الفرضية لأنها تختلف عن مستوى المحدد وهو ٥% .

٦.٤.٢ أختبار kstest

وهو اختبار kolmogorov Smirnov والذي يستخدم لاختبار توزيع معين ولعينة واحدة.

فإذا كانت قيمة h = 1 فبإمكانك أن ترفض الفرضية ضمن المستوى α وإذا كانت قيمة h = 0 فلا يمكنك أن ترفض الفرضية ضمن المستوى α.

مثالاً على ذلك نولد عينة بقيم من ٢- إلى ٤ وبزيادة واحد في كل مرحلة حيث تتولد سبعة قيم وكما يلي:

```
>> X = - 2 : 1 : 4

X =  -2   -1   0   1   2   3   4
```

بعدها نطبق أختبار kstest على المتغير X وكما يلي:

```
[ h , p , k , c ]  =  kstest  ( X , [] ,  0.05 , 0 )

h  =  0
p  = 0.13632
k  =  0.41277
c  =  0.48342
```

وبما أن h = 0 فأنه لا يمكنك أن ترفض الفرضية أي أن العينة مأخوذة من التوزيع الطبيعي.

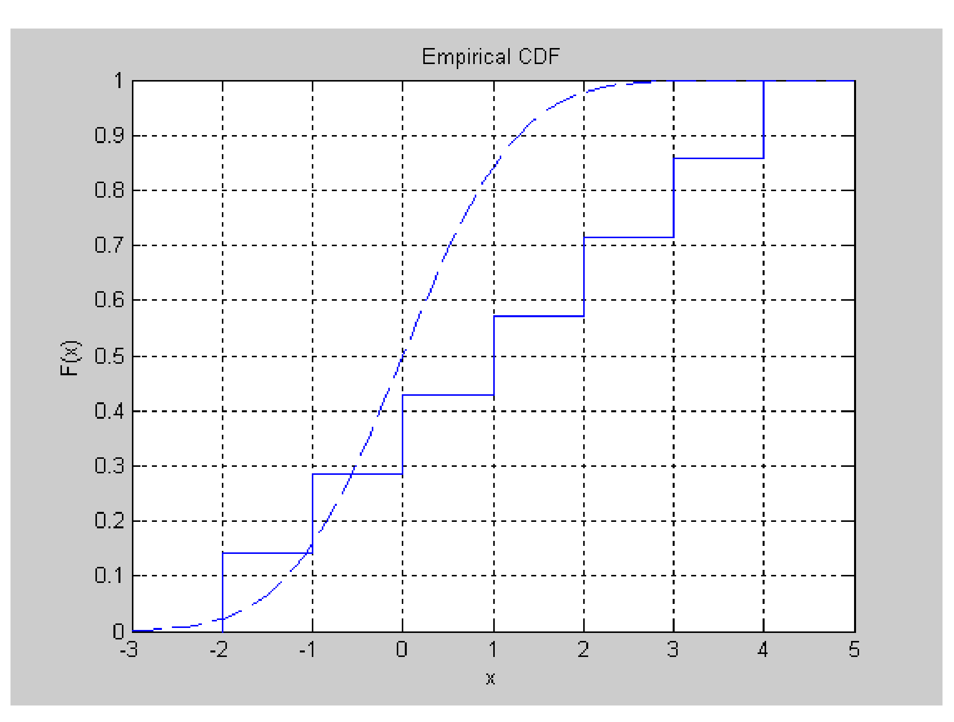
ولفهم الاختبار نولد عينة من نوع التوزيع التراكمي Empirical cumulative distribution وعليه نولد عينة مكونة من القيم من ٣- إلى ٥ وبزيادة ٠.١ ثم نرسم المتغير X باستخدام الدالة cdfplot وبعدها نرسم المتغير XX باستخدام الدالة plot كما موضحة في الشكل ٦.٩ حيث يمكن ملاحظة الفرق بين المحنيين وتكتب الاوامر كما يلي:

```
XX  = -3 : 0.1:5

xx =

   -3.0000   -2.9000   -2.8000   -2.7000   -2.6000
   -2.5000   -2.4000   -2.3000   -2.2000   -2.1000
   -2.0000   -1.9000   -1.8000   -1.7000   -1.6000
   -1.5000   -1.4000   -1.3000   -1.2000   -1.1000
   -1.0000   -0.9000   -0.8000   -0.7000   -0.6000
   -0.5000   -0.4000   -0.3000   -0.2000   -0.1000
         0    0.1000    0.2000    0.3000    0.4000
    0.5000    0.6000    0.7000    0.8000    0.9000
    1.0000    1.1000    1.2000    1.3000    1.4000
    1.5000    1.6000    1.7000    1.8000    1.9000
    2.0000    2.1000    2.2000    2.3000    2.4000
    2.5000    2.6000    2.7000    2.8000    2.9000
    3.0000    3.1000    3.2000    3.3000    3.4000
    3.5000    3.6000    3.7000    3.8000    3.9000
    4.0000    4.1000    4.2000    4.3000    4.4000
    4.5000    4.6000    4.7000    4.8000    4.9000
    5.0000

>> cdf plot   (X)
>> hold on
>> plot ( XX, normcdf (XX ) , ‘b - -‘)
```



شكل ٦.٩ التوزيع لعينتين مختلفتين كما في المثال السابق بأستخدام cdfplot

الشكل ٦.٩ يوضح أن أعلى قيمة هي ٠.٤١٢٧٧ تحدث عند تكون قيمة X = 0.1 ويمكن ملاحظة ذلك مـن المنحنـى حيـث تظهـر لـه قيمـة ٣/٧ وأنها تسـاوي ٠.٤١٢٧٧. وكذلك بالامكـان إجـراء الاختبـار مـن وجهـة نظـر غـير متسـاوية وتكـون صحيحة وذلك بأستخدام الدالة kstest وكما يلي:

```
>> [h,p,k] = kstest(x, [], .05, -1)

h =

    0

p =

   0.0682

k =

   0.4128
```

هذا الاختبار هو مشابه للسابق فيما عـدا أن قيمـة P تكـون قليلـة لأنهـا تعطـي جانب واحد ولتطبيق الاختبار على الجانب الآخر كما يلي:

```
>> [h,p,k] = kstest(x,[],0.05,1)
h =

   0

p =
   0.7753
k =
   0.1271
```

الان نولد عينة عشوائية بمئة قيمة من توزيع Weibull وكما يلي:

```
>> x = wblrnd(1, 2, 100, 1)
```

```
x =
0.2262   1.2103   0.7067   0.8495   0.3392   0.5212
0.8856   1.9974   0.4435   0.9002   0.6967   0.4830
0.2853   0.5509   1.3175   0.9498   0.2583   0.2945
0.9439   0.3353   1.6880   1.0206   0.4548   2.1492
1.4050   1.2632   1.2712   0.7103   1.1407   1.2710
2.0449   0.5403   0.8997   0.2657   0.8738   0.9331
0.4086   0.8025   1.2634   0.6303   0.4202   1.9825
0.6195   0.9844   0.4291   0.8292   0.5859   0.9201
1.0903   1.2894   1.2817   0.6184   1.0931   0.7830
1.3753   0.5997   0.9858   0.3883   0.3978   0.7222
0.8367   0.3250   0.4432   0.6623   0.4482   0.6443
1.0359   1.1130   1.0370   0.7920   0.5645   1.0833
0.4197   0.7520   0.9966   0.5939   0.7772   0.9000
0.6037   0.6899   0.4792   0.2100   0.8056   0.3573
1.3247   0.1430   1.1419   1.1735   0.3643   0.5520
1.4111   2.1079   0.3349   1.2703   1.0992   0.6429
1.1213   0.8699   1.6543   0.1083
```

ثم نختبر العينة مع توزيع Weibull حيث نحصل على الناتج صفر أي أن العينة مطابقة وكما يلي:

```
>> kstest(x, [x wblcdf(x, 1, 2)])

ans =

    0
```

ثم نختبر العينة مع توزيع exponential حيث نحصل على الناتج واحد أي أن العينة غير مطابقة وكما يلي:

```
>> kstest(x, [x expcdf(x, 1)])

ans =

    1
```

٦.٤.٣ أختبار kstest2

وهو اختبار kolmogorv Smirnov ويستخدم لاختبار مقارنة التوزيع لعينتين. فإذا كانت قيمة h = 1 يعني نرفض الفرضية على كون التوزيع في العينتين نفسه، أما إذا كانت قيمة h = 0 يعني عدم رفض الفرضية أي أن التوزيع في العينتين نفسه حيث أن القياس يكون على أساس نسبة المستوى ٥%.

ولتحقيق هذا الاختبار نأخذ عينتين حيث تمثل العينة الأولى تغيير للقيم من ١- إلى ٥ وبزيادة ١ في كل خطوة حيث تتكون سبعة عناصر في المتغير وتكتب كما يلي:

```
>> x = -1:1:5

x =

   -1    0    1    2    3    4    5
```

أما العينة الثانية فتتكون من عشرين قيمة عشوائية وتولد باستخدام الدالة randn وكما يلي:

```
>> y = randn(20,1)

y =

   0.2944   -1.3362    0.7143    1.6236   -0.6918

   0.8580    1.2540   -1.5937   -1.4410    0.5711

  -0.3999    0.6900    0.8156    0.7119    1.2902

   0.6686    1.1908   -1.2025   -0.0198   -0.1567
```

ننفذ الاختبار باستخدام الدالة kstest2 لمقارنة التوزيع في العينتين وكما يلي:

```
>> [h,p,k] = kstest2(x,y)

h =

    1

p =

   0.0403

k =

   0.5714
```

بما أن h=1 يعني ذلك رفض الفرضية أي أن هناك فرق بين العينتين وكذلك فأن قيمة P = 4% يعني أنها ضمن حدود إلى ٥% وهي القيمة التي يجري عليها القياس.

ويمكن التحقق من الفرق بين العينتين باستخدام الدالة cdfplot لرسم العينتين وكما موضحة في الشكل ٦.١٠ وتستخدام الدالة findobj لإيجاد الأجسام بقيم معينة وتكتب الاوامر كما يلي:

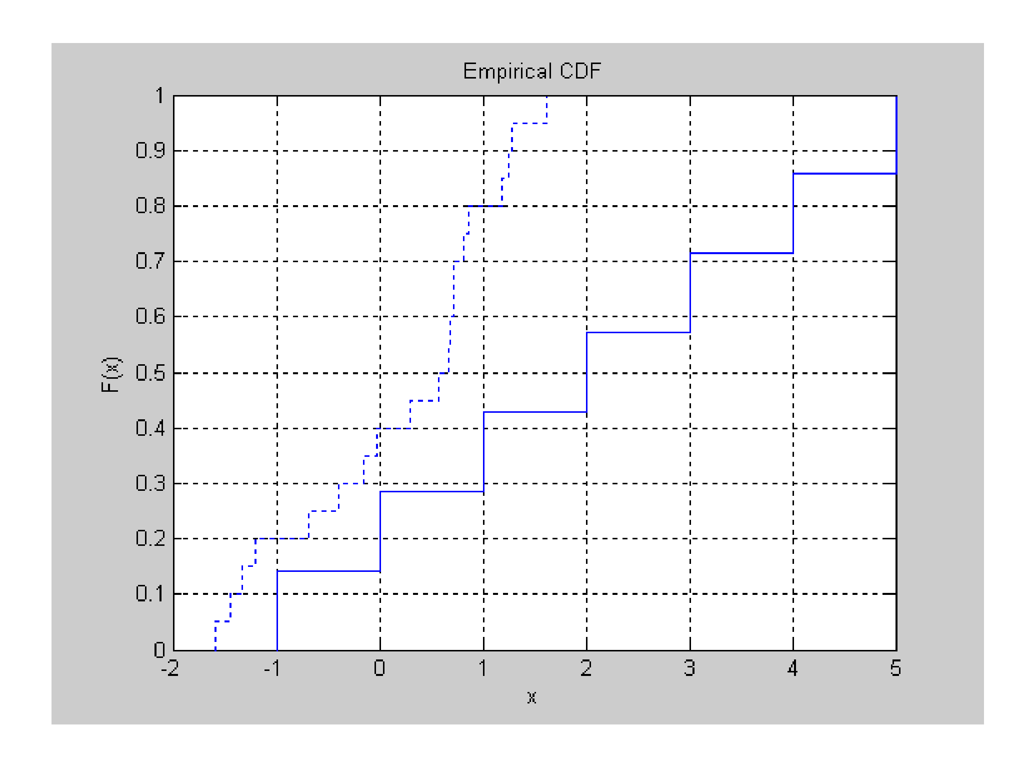
```
>> cdfplot(x)

>> hold on

>> cdfplot(y)

>> h = findobj(gca,'type','line');

>> set(h(1),'linestyle',':','color','b')
```



شكل ٦.١٠ الفرق بين العينتين في المثال السابق بأستخدام الدالة cdfplot

٦.٤.٤ أختبار lillietest

وهو اختبار Lilliefors ويستخدم لإيجاد جودة أو إمكانية التوافق مع التوزيع الطبيعي. فإذا كانت قيمة h=1 يعني بإمكانك رفض الفرضية على كون العينة ذات توزيع طبيعي وإذا كانت قيمة h = 0 لا يمكنك رفض الفرضية علماً بأن القيمة التي يجري عليها القياس بمستوى ٥%.

مثال على ذلك نحمل الملف carsmall والخاصة بالمعلومات عن السيارات الصغيرة ومن هذه المعلومات الاوزان وذلك بأستخدام الدالة load وكما يلي:

```
>> load carsmall
>> Weight
```

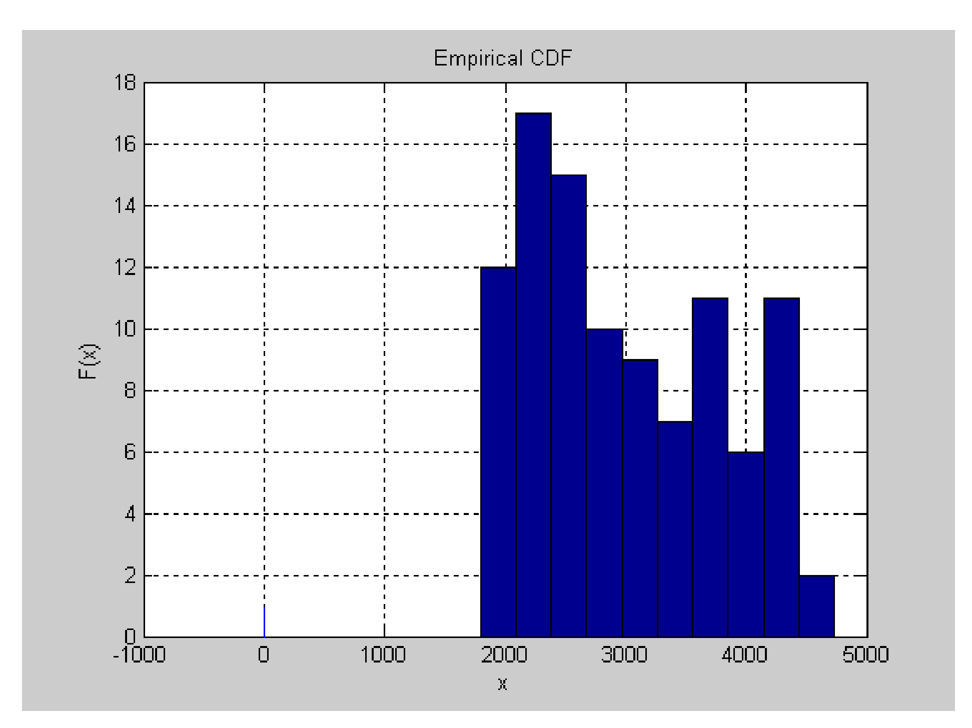
```
Weight =
3504    3693    3436    3433    3449    4341
4354    4312    4425    3850    3090    4142
4034    4166    3850    3563    3609    3353
3761    3086    2372    2833    2774    2587
2130    1835    2672    2430    2375    2234
2648    4615    4376    4382    4732    2464
2220    2572    2255    2202    4215    4190
3962    4215    3233    3353    3012    3085
2035    2164    1937    1795    3651    3574
3645    3193    1825    1990    2155    2565
3150    3940    3270    2930    3820    4380
4055    3870    3755    2605    2640    2395
2575    2525    2735    2865    3035    1980
2025    1970    2125    2125    2160    2205
2245    1965    1965    1995    2945    3015
2585    2835    2665    2370    2950    2790
2130    2295    2625    2720
```

نطبق الاختبار باستخدام الدالة Lillietest على أوزان السيارات Weight وكما يلي:

```
>> [h p l c] = lillietest(Weight);
>> [h p l c]
ans =
   1.0000    0.0232    0.1032    0.0886
```

ومن النتائج يلاحظ أن القيمة الإحصائية للاختبـار ٠.١٠٣٢ أكـبر مـن القيمـة الحرجة ٠.٠٨٨٦ لذلك ترفض الفرضية وهذا واضح حيث أن قيمة h=1 وكذلك فـأن قيمة p = ٢% تقريباً. ولإيضـاح ذلـك بالإمكـان اسـتخدام الدالـة hist لعـرض نتـائج توزيع أوزان السيارات وكما موضحة في الشكل ٦.١١ وتكتب الاوامر كما يلي:

```
>> hist(Weight)
```



شكل ٦.١١ عرض البيانات في المثال السابق بأستخدام المخطط التكراري

وبالأمكان تطبيق اللوغارتم على عينة الاوزان وذلك بأستخدام الدالة log وكما يلي:

```
>> [h p l c] = lillietest(log(Weight))

ans =

        0      0.13481      0.077924      0.0886
```

٦.٤.٥ أختبار ranksum

وهو اختبار Wilcoxon ويستخدم لإيجاد الوسيط المتساوي. فإذا كانت قيمة h=0 يعني لا يمكنك رفض الفرضية أي أن الوسيط للعينتين متساوي أما إذا كانت h=1 فهذا يعني رفض الفرضية أي أن الوسيط للعينتين غير متساوي.

كمثال على ذلك نولد عينتين، بأحجام مختلفة، العينة الأولى عبارة عن عشرة قيم عشوائية من التوزيع الاعتيادي المستمر (Continuous Unform distribution) بأستخدام الدالة unifrnd وكما يلي:

```
>> x = unifrnd(0,1,10,1)

x =

   0.5828    0.4235    0.5155    0.3340    0.4329

   0.2259    0.5798    0.7604    0.5298    0.6405
```

أما العينة الثانية فهي عبارة عن خمسة عشر قيمة عشوائية من التوزيع المستمر وتولد بأستخدام الدالة unifrnd وكما يلي:

```
>> y = unifrnd(.25,1.25,15,1)

y =

   0.4591    0.6298    1.0333    0.9308    0.7111

   0.8178    1.0442    0.3092    0.8529    0.3003

   0.6654    0.5550    1.1244    0.2650    1.0180
```

بعد ذلك نطبق اختبار ranksum على العينتين وكما يلي:

```
>> [p,h] = ranksum(x,y,0.05)

p =
   0.0631
h =
    0
```

هذا يعني أنه لايمكن رفض الفرضية أي أن التوزيع هنا يكون مماثل للعينتين وذلك لأن h=0.

٦.٤.٦ أختبار signrank

وهو اختبار Wilcoxon ويستخدم لإيجاد الوسيط الصفري.

فإذا كانت قيمة h=0 يعني أن الوسيط للعينتين صفر أي لا يمكن رفض الفرضية أما إذا كانت h=1 فيعني رفض الفرضية.

كمثال على ذلك نولد عينتين الأولى عبارة عن عشرة قيم للعينة عشوائية من التوزيع الطبيعي اللوغارتمي بأستخدام الدالة lognrnd وكما يلي:

```
>> before = lognrnd(2,.25,10,1)

before =

   4.9480    7.8800    5.6739   10.5253    6.0420

   8.4333    7.8055    5.8681    4.2945    7.2805
```

والعينة الثانية عبارة عن عشرة قيم للعينة الأولى مضاف إليها عينة عشوائية لتوزيع أخر وتكون كما يلي:

```
>> after = before + trnd(2,10,1)

after =

   3.7102    8.8070    6.0312   28.7266    6.6107

   7.9394    8.1204    4.9686    4.2712    7.2396
```

بعد ذلك نطبق اختبار signrank على العينتين وكما يلي:

```
>> [p,h] = signrank(before,after,0.05)

p =

   0.7695

h =

    0
```

هذا يعني أنه لا يمكن رفض الفرضية لان قيمة h=0 .

٦.٤.٧ أختبار ttest

وهو اختبار الفرضية لإيجاد معدل العينة. فإذا كانت h=0 فأن الفرضية لا يمكن رفضها أي أن المعدل هو صفر أما إذا كانت h=1 فأنه يمكن رفض الفرضية على أساس بنسبة المستوى ٥%.
كمثال على ذلك نولد عينة بمئة قيمة وتكون عشوائية من التوزيع الطبيعي وكما يلي:

```
>> x = normrnd(0,1,1,100)
x =
   0.5689  -0.2556  -0.3775  -0.2959  -1.4751
  -0.2340   0.1184   0.3148   1.4435  -0.3510
   0.6232   0.7990   0.9409  -0.9921   0.2120
   0.2379  -1.0078  -0.7420   1.0823  -0.1315
   0.3899   0.0880  -0.6355  -0.5596   0.4437
  -0.9499   0.7812   0.5690  -0.8217  -0.2656
  -1.1878  -2.2023   0.9863  -0.5186   0.3274
   0.2341   0.0215  -1.0039  -0.9471  -0.3744
  -1.1859  -1.0559   1.4725   0.0557  -1.2173
  -0.0412  -1.1283  -1.3493  -0.2611   0.9535
   0.1286   0.6565  -1.1678  -0.4606  -0.2624
  -1.2132  -1.3194   0.9312   0.0112  -0.6451
   0.8057   0.2316  -0.9898   1.3396   0.2895
   1.4789   1.1380  -0.6841  -1.2919  -0.0729
  -0.3306  -0.8436   0.4978   1.4885  -0.5465
  -0.8468  -0.2463   0.6630  -0.8542  -1.2013
  -0.1199  -0.0653   0.4853  -0.5955  -0.1497
  -0.4348  -0.0793   1.5352  -0.6065  -1.3474
   0.4694  -0.9036   0.0359  -0.6275   0.5354
   0.5529  -0.2037  -2.0543   0.1326   1.5929
```

بعدها تجري اختبار ttest عليها وكما يلي:

```
>> [h,p,ci] = ttest(x,0)
h =
    0
p =
   0.0963
ci =
  -0.3070    0.0256
```

والنتيجة h=0 يني لا يمكن رفض الفرضية وأن قيمة p تساوي 0.0963 يعني القيمة مقبولة.

٦.٤.٨ أختبار ttest٢

وهو اختيار الفرضية للفرق في المعدل لعينتين. فإذا كانت h=1 فبإمكانك رفض الفرضية يعني المعدل متساوي للعينتين بنسبة ٥% أما إذا كانت قيمة h=0 يعني أن هناك فرق في المعدل.

كمثال على ذلك نأخذ عينتين الأولى تكون قيمة لتوزيع طبيعي ولمعدل صفر وانحراف معياري يساوي واحد وكما يلي:

```
>> x = normrnd(0,1,100,1)
```

```
x =
1.0184    -1.5804   -0.0787   -0.6817   -1.0246
-1.2344   0.2888    -0.4293   0.0558    -0.3679
-0.4650   0.3710    0.7283    2.1122    -1.3573
-1.0226   1.0378    -0.3898   -1.3813   0.3155
1.5532    0.7079    1.9574    0.5045    1.8645
-0.3398   -1.1398   -0.2111   1.1902    -1.1162
0.6353    -0.6014   0.5512    -1.0998   0.0860
-2.0046   -0.4931   0.4620    -0.3210   1.2366
-0.6313   -2.3252   -1.2316   1.0556    -0.1132
0.3792    0.9442    -2.1204   -0.6447   -0.7043
-1.0181   -0.1821   1.5210    -0.0384   1.2274
-0.6962   0.0075    -0.7829   0.5869    -0.2512
0.4801    0.6682    -0.0783   0.8892    2.3093
0.5246    -0.0118   0.9131    0.0559    -1.1071
0.4855    -0.0050   -0.2762   1.2765    1.8634
-0.5226   0.1034    -0.8076   0.6804    -2.3646
0.9901    0.2189    0.2617    1.2134    -0.2747
-0.1331   -1.2705   -1.6636   -0.7036   0.2809
-0.5412   -1.3335   1.0727    -0.7121   -0.0113
-0.0008   -0.2494   0.3966    -0.2640   -1.6640
```

والعينة الثانية تكون قيمة لتوزيـع طبيعـي ولمعـدل ٠.٥ وانحـراف معيـاري يسـاوي واحد وكما يلي:

```
>> y = normrnd(0.5,1,100,1)
```

```
y =
-0.5290    0.7431    -0.7566    0.1528    -0.4414
 -0.6746   -0.5211    0.0983     0.6737    0.3839
1.5641     0.2546    -1.0175    0.5097     0.5714
0.8165     0.9998    1.7781     -0.0478   0.7608
  0.4868   -0.0803   2.6363     0.2424    -0.9095
2.2701     0.8255    -0.6190    1.1204     1.7698
-0.3960    0.6352    0.3610     -0.6634    1.6837
0.4846     1.0362    -0.2164    -0.1556    0.8144
0.6068     2.3482    0.2249     2.7126     2.0085
-1.4451    -1.1805   -0.0735    0.3142     0.5089
1.3369     -0.2223   -0.2215    0.2988     0.4795
0.7789     1.5583    1.1217     -1.2506    1.1973
1.3115     1.1363    1.8101     0.8271     -0.1730
0.3507     -1.9490   0.9733     0.6169     -0.0911
-0.1547    -0.5807   0.4523     0.8793     0.1696
0.0001     0.4640    0.3252     -0.4573    1.7925
0.9409     1.7809    0.0023     -0.6187    1.3076
0.5412     -0.2562   0.4109     -1.5089    1.5839
-0.4812    -0.1885   1.8395     -0.4092    0.0871
```

ثم بعد ذلك نطبق اختبار ttest على العينتين وكما يلي:

```
>> [h,significance,ci] = ttest2(x,y)
h =
    1
significance =
  7.6592e-004
ci =
   -0.7406   -0.1987
```

النتيجة توضح أن h=1 وهذا يعني ترفض الفرضية وبما أن القيمة المؤثرة تساوي 7.6592e-004 فالقيمة المستحصلة هي مؤثرة أي ٩٥%.

٦.٤.٩ أختبار ztest

وهو اختيار الفرضية للمعدل لعينة معينة إذا كان التباين معروف. فإذا كانت قيمة h=1 فبإمكانك رفض الفرضية أما إذا كانت قيمة h=0 فتقبل الفرضية. ولتحقيق ذلك نولد عينة بمئة قيمة لتوزيع طبيعي وبمعدل يساوي صفر وانحراف معياري يساوي واحد وكما يلي:

```
>> x = normrnd(0,1,100,1)
```

```
x =
-0.4326   -1.6656   0.1253   0.2877   -1.1465   1.1909
1.1892    -0.0376   0.3273   0.1746   -0.1867   0.7258
-0.5883   2.1832    -0.1364  0.1139   1.0668    0.0593
-0.0956   -0.8323   0.2944   -1.3362  0.7143    1.6236
-0.6918   0.8580    1.2540   -1.5937  -1.4410   0.5711
-0.3999   0.6900    0.8156   0.7119   1.2902    0.6686
1.1908    -1.2025   -0.0198  -0.1567  -1.6041   0.2573
-1.0565   1.4151    -0.8051  0.5287   0.2193    -0.9219
-2.1707   -0.0592   -1.0106  0.6145   0.5077    1.6924
0.5913    -0.6436   0.3803   -1.0091  -0.0195   -0.0482
0.0000    -0.3179   1.0950   -1.8740  0.4282    0.8956
0.7310    0.5779    0.0403   0.6771   0.5689    -0.2556
-0.3775   -0.2959   -1.4751  -0.2340  0.1184    0.3148
1.4435    -0.3510   0.6232   0.7990   0.9409    -0.9921
0.2120    0.2379    -1.0078  -0.7420  1.0823    -0.1315
0.3899    0.0880    -0.6355  -0.5596  0.4437    -0.9499
0.7812    0.5690    -0.8217  -0.2656
```

نحسب المعدل حيث تكون القيمة قليلة ومقارب إلى الصفر ويكتب الامر كما يلي:

```
>> m = mean(x)
m =
   0.0479
```

بعدها نطبق اختبار ztest على العينة وكما يلي:

```
>> [h,sig,ci] = ztest(x,0,1)
h =
    0
sig =
   0.6317
ci =
  -0.1481    0.2439
```

النتيجة h = 0 تعني أنه لا يمكن رفض الفرضية وأن القيمة المؤثرة 0.6317 والتي تعني أنه ٦٣ قيمة من كل مئة قيمة تكون مؤثرة.

الفصل السابع

تطبيقات حزمة MATLAB

في النماذج الخطية / تحليل التباين

Matlab Package Applications

on Linear Models (ANOVA)

٧.٨ تطبيق على تحليل التباين متعدد الطرق لمجموعة كبيرة

Example of large group N-way Analysis of Variance

٧.٩ تحليل التباين بالتأثيرات العشوائية

ANOVA with Random Effects

٧.٩.١ تجهيز النموذج **Setting Up the Model**

٧.٩.٢ موائمة نموذج التأثيرات العشوائية

Fitting a Random Effects Model

٧.٩.٣ احصاء F للنماذج بالتأثيرات العشوائية

Fitting a Random Effects Mode

٧.٩.٤ مركبات التباين **Variance Components**

٧.١ مقدمة Introduction

سوف نتطرق في هذا الفصل الى النماذج الخطية وتحليلات التبـاين بـالطرق المختلفة وكـذلك نسـتعرض أمثلـة موجـودة ضـمن تطبيقـات الحزمـة. تعتبرالـنماذج الخطيـة احـدى التطبيقـات المهمـة في الحزمـة الاحصـائية ضـمن تطبيقـات حزمـة MATLAB وهي تمثل العلاقة بين المتغير ذو الاستجابة المستمرة وواحد أو أكثر مـن المتغيرات التي يتم التنبأ بها وتكون العلاقة كما يلي :

$$y = x\beta + \varepsilon$$

حيث ان :

y يمثل صف من متغير الاستجابة .

x يمثل مجموعة n*p تحسبت من الاستجابات

β يمثل صفوف من المتغيرات

ε يمثل صف n*1 من التغير العشوائية والغير معتمدة على بعضها .

حزمة MATLAB تستخدم هذه العلاقـة للنمـوذج الخطـي لحـل التنـوع أو أرتـداد معينة وتحليل مسائل التبـاين . وسـوف نسـتعرض مجموعـة مـن الفقـرات الخاصـة بالنماذج الخطية وتشمل:

١. تحليل التباين ذو طريق واحد
٢. تحليل التباين ذو طريقين
٣. تحليل التباين ذو عدة طرق
٤. تحليل التباين بالتأثيرات العشوائية

٧.٢ تحليل التباين ذو طريق واحد

One-way Analysis Of Variance (ANOVA)

تحليل التباين ذو طريق واحد يهدف الى ايجاد فيما إذا كانـت البيانـات مـن بين المجاميع العديدة تحتـوي معـدل مشـترك، وهـذا يعنـي حسـاب فـيما إذا كانـت المجاميع هي في الحقيقة مختلفـة في الخـواص المحسـوبة. تحليـل التبـاين ذو طريـق واحد هي حالة خاصة وبسيطة للنموذج الخطي وعلاقتها تكون كما يلي :

$$Yij = \alpha j + \varepsilon ij$$

حيث ان :

Yij عبارة عن مصفوفه فيها كل عمود يمثل مجموعة مختلفة.

αj عبارة عن مصفوفه فيها ألاعمدة تمثل معدلات المجاميع.

εij عبارة عن مصفوفه تمثل التغيرات العشوائية.

هذه العلاقة تمثل النموذج الخطي بأفتراض ان الاعمدة Y عبارة عن ثابت مضافا اليه التغير العشوائي.

٧.٣ تطبيق على تحليل التباين ذو طريق واحد — Example of ANOVA

سوف نستعرض هنا مثال على تحليل التباين ذو طريق واحد حيث نأخذ نموذج مخزون في الحزمة وهو حساب عدد البكتيريا في ارساليات الحليب حيث أن أعمدة المصفوفه تمثل الارساليات المختلفة اما الصفوف فتمثل حساب البكتيريا لكل علبه ويكون الاختيار بشكل عشوائي من الارسالية والتي يتم من خلالها معرفة الارساليات التي تحتوي على كمية أكبر من البكتيريا بالنسبة الى الارساليات الاخرى .

نحمـل البيانـات الخاصـة باحصـائيات البكتيريـا والموجـود في الملـف hogg باستخدام الدالة lood ثم نعرض محتوى الملف وكما يلي :

lood hogg >>

hogg>>

```
hogg =
24   14   11   7   19
15   7    9    7   24
21   12   7    4   19
27   17   13   7   15
33   14   12   12  10
23   16   18   18  20
```

نحسب ANOVA على البيانات السابقة وكما يلي :

[P, tbl , stats] = anova1 (hogg) > >

P =

1.1971e-004

```
tbl =
'Source'    'SS'    'df'    'MS'    'F'    'Prob>F'    'Columns'
[803.0000]  [ 4]  [200.7500] [9.0076] [1.1971e-004]
'Error'      [557.1667]     [25]      [ 22.2867]      []      []
'Total'      [1.3602e+003]  [29]      []              []      []
```

stats =

gnames: [5x1 char]

n: [6 6 6 6 6]

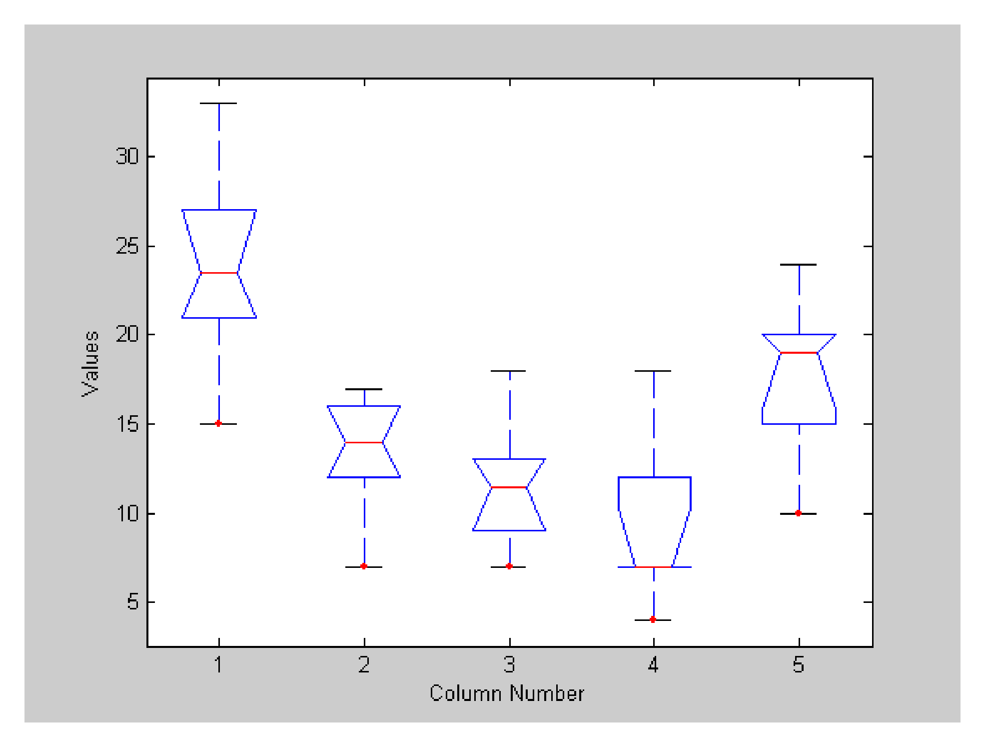
source: 'anova1'

means: [23.8333 13.3333 11.6667 9.1667 17.8333]

df: 25

s: 4.7209

جدول ANOVA القياسي يحتوي على خمسة اعمدة وهي من اليسار مجموع المربعات ، درجات الحرية ، معدل المربعات ، احصاء F وقيمة P وهي كذلك ممثلة في الشكل ٧.١.



شكل ٧.١ مخطط لحساب كمية البكتريا في الملف hogg

يمكن استخدام احصاء f لعمل اختبارالفرضية لايجادفيما إذا كان حساب البكتريا نفسها، فالدالة anoval ترجع قيمة p من اختبارالفرضية وفي هذه الحالة فان قيمة p هي بحدود 0.0001 وهي قيمةقليلة. وهذا هو مؤشر قوي والذي فيه يتم حساب البكتيريا من عدة ارساليات وهي غيرمتساوية. أحصاء f يعطي حالةقصوى من التطرف حيث ان f يحدث بنسبة واحدمن ١٠٠٠٠. قيمة p الراجعةمن anoval تعتمدعلى فرضية

التوزيعات العشوائية فأذا كانت قيمة p صحيحة فهذه التوزيعات تحتاج لتكون غير معتمدة وموزعة طبيعيا وتحتوي على تباين ثابت.

فلو تم توظيف هذه الدالة بالشكل الصحيح لكان من السهل اعطاء التحليل الصحيح وعلى ضوء ذلك تهدف النتائج الى المساعدة في اختيار القرار الصحيح وفي تحديد ذلك القرار .

هنا محاولة لتمثيل سلسلة من اختبارات t ،واحد لكل زوج من قيم المعدلات ولكن هذا الاجراء فيه خطر. اختبار t يحسب احصاء t ومقارنته للقيمة الحرجة، حيث أن القيمة الحرجة اختيرت لذلك عندما تكون قيم المعدلات يكون نفسها وان احتمالية كون احصاء t سوف تزيد القيمة الحرجة حيث تكون صغيرة بحدود ٥%. عندما تكون قيم المعدلات مختلفة فان الاحتمالية التي تكون فيها الاحصاء يزيد على القيمة الحرجة تكون اكبر.

في المثال السابق توجد خمسة قيم للمعدلات لذلك توجد عشرة أزواج من قيم المعدلات للمقارنة. فيما إذا كانت جميع قيم المعدلات نفسها وكذلك كانت ٥% غير صحيح نستنتج أنه يوجد اختلاف في زوج واحد ثم الاحتمالية تعمل على الاقل استنتاج واحد غير صحيح بين عشرة ازواج يكون اكبر من ٥% .

يمكن اختيار تعدد المقارنات باستخدام داله multcompare وتجهيزها مع حالات anova1 وكما يلي :

```
[ c,m] = multcompare (stats)  >>
```

المخرجات الأولى من داله multcompare لها صف واحد لكل زوج من المجاميع يتضمن الفرق لقيم مجموعة المعدلات ومدى التحقق للمجموعة ، فمثلا الصف الثاني يتضمن القيم .

```
1.0000   2.0000    2.4953   10.5000   18.5047
1.0000   3.0000    4.1619   12.1667   20.1714
1.0000   4.0000    6.6619   14.6667   22.6714
1.0000   5.0000   -2.0047    6.0000   14.0047
```

2.0000	3.0000	-6.3381	1.6667	9.6714
2.0000	4.0000	-3.8381	4.1667	12.1714
2.0000	5.0000	-12.5047	-4.5000	3.5047
3.0000	4.0000	-5.5047	2.5000	10.5047
3.0000	5.0000	-14.1714	-6.1667	1.8381
4.0000	5.0000	-16.6714	-8.6667	-0.6619

m =

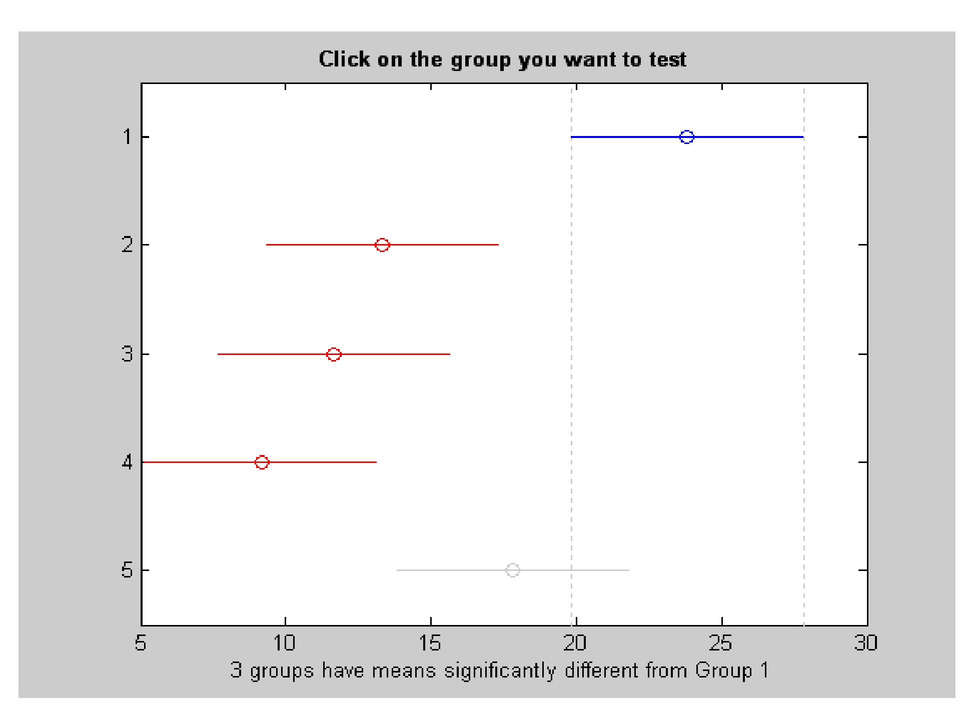
23.8333	1.9273
13.3333	1.9273
11.6667	1.9273
9.1667	1.9273
17.8333	1.9273

الاخراج الأول من داله multcompare لهـا صـف واحـد لكـل زوج مـن المجـاميع ويتضمن الفرق لقيم مجموعة المعدلات ومدى التحقـق للمجموعـة ، فمـثلا الصـف الثاني يتضمن القيم .

1.0000 3.0000 4.1619 12.1667 20.1714

وهذا يعني ان المعدل للمجموعـة ١ ناقصـا مجموعـة ٣ خمنـت لتكـون ٢٠.١٧١٤ و ٩٥% لمدى التحقق للفرق هـو [20.1714 ، ٤.١٦١٩] وهـذا المـدى لا يحتـوي عـلى صفر لذلك يمكن ان تستنتج بان قيم المعدلات للمجموعة ١ ، ٣ تكون مختلفة.

الأخراج الثاني يحتوي على المعدل والخطأ القياسي لكل مجموعة ، ومن السهل ايضاح الفرق بين قيم المعدلات بنظرة على الشكل 7.2 .



شكل ٧.٢ الفرق بين قيم المعدلات في المثال السابق بأستخدام الدالة multcompare

فلتوظيـف دالـه multcompare بشـكل صـحيح يمكـن تحديـد مقاومـات متعددة وعلى ضوئها اختبار المجاميع المناسبة وكـون المعـدلات متشـابهة أو مختلفـة حيث يمكن توظيفها في مجالات مختلفة لنظم المعلومات .

٧.٣ تحليل التباين ذو طريقين(Two-Way Analysis of Variance (ANOVA

الهدف من تحليل التبـاين ذو أتجـاهين هـو لايجـادفيما إذا كانـت البيانـات مـن عدةمجاميع لها نفس قيمة المعدل. يختلف تحيليل التبـاين ذوطـريقنين عـن ذوالطريـق الواحدحيث تكون المجاميع في ANOVA ذوالطريقتين لها صنفين لتعريف الخـواص بـدلاً من صنف واحد. لوفرضنا أن شركة سيارات لهامصنعين وكل مصنع ينـتج نفـس المـوديلات الثلاثةالمقرةللسيارة حيث أنه من المجدي أن يطرح سؤال فيما إذا كانت المسافة المقطوعة

للتر الواحد من الوقود في السيارات تتغير لنفس الموديلات الثلاث. هنا يبرز لدينا مؤشرين هما المصنع والموديل لتوضيح الاختلاف في المسافة المقطوعة بالأميال لكل لتر من الوقود.

يجب أن يكون هناك فرق عام في المسافة المقطوعة بالأميال لكل لتر من الوقود من خلال الاختلاف في طرق الانتاج بين المصانع. وكذلك من المحتمل وجود اختلاف في المسافة المقطوعة بالأميال لكل لتر للموديلات المختلفة بسبب الاختلاف في مواصفات التصميم وهذه التأثيرات تدعى التأثيرات المضافة. كذلك المصنع ربما ينتج موديل واحد تكون فيه المسافة المقطوعة بالأميال لكل لتر وقود أكبر وهذا بسبب كون الخط الانتاجي ذو أهمية أكبر ولكنه لا يكون مختلف عن المصنع الآخر للموديلات الأخرى وهذا التأثير يسمى بالتأثير التفاعلي.

فلو تم توظيف هذه الدالة وادخال كافة المؤثرات التي تدخل في العملية الانتاجية وبرمجة ذلك في خوارزمية أو نظام معلومات متكامل لكان بالامكان التوصيل إلى تكييف العمل والحصول على نتائج أفضل وهذا النظام المعلوماتي يساعد بشكل كبير على أتحاذ القرار الصائب وتعديل كافة الاخطاء الناتجة عن التصنيع أو المؤثرات الاخرى.

تحليل التباين ذو طريقتين هو حالة خاصة للنموذج الخطي وتمثل بالمعادلة التالية:

$$y_{ijk} = \mu + \alpha_j + \beta_i + \gamma_{ij} + \varepsilon_{ijk}$$

حيث أن هذه المعادلة لو طبقت على المثال السابق والخاص بالمسافة المقطوعة للسيارات لكانت المتغيرات كما يلي:

y_{ijk} :تمثل مصفوفة ثابتة عامة للمسافة المقطوعة بالأميال لكل لتر وقود.

α_j : تمثل مصفوفة فيها الأعمدة تمثل ألانحراف لكل سيارة حسب المسافة المقطوعة بالأميال لكل لتر وقود.

μ : تمثل معدل المسافة المقطوعة بالأميال لكل لتر وقود.

γ_{ij} : تمثل مصفوفة التفاعلات:

ε_{ijk} : تمثل مصفوفة التوزيعات العشوائية.

٧.٥ تطبيق على تحليل التباين ذو طريقين

Example of Two-Way Analysis of Variance

وكتطبيـق عـلى تحليـل التبـاين نأخـذ مثـال لحسـاب تـأثير موديـل السـيارة والمصنع على المسافة المقطوعة بالأميال لكـل لـتر مـن الوقـود للسـيارات وفي البدايـة نحمل الملف mileage باستخدام الدالة load ثم تعرض محتويات الملف بكتابة أسم المتغير mileage وكما يلي:

```
>> load mileage

>> mileage

mileage =

   33.3000   34.5000   37.4000

   33.4000   34.8000   36.8000

   32.9000   33.8000   37.6000

   32.6000   33.4000   36.6000

   32.5000   33.7000   37.0000

   33.0000   33.9000   36.7000
```

هنا عـدد السـيارات المسـتخدمة في الاختبـار هـي ثلاثـة وكـل مـنهم خضـعت لسـتة أختبارات. وبعد ذلك نحسب المأثرات باستخدام الدالة anova2 وكما يلي:

```
>> cars = 3

>> [p,tb1, stats] = anova2 (mileage, cars)

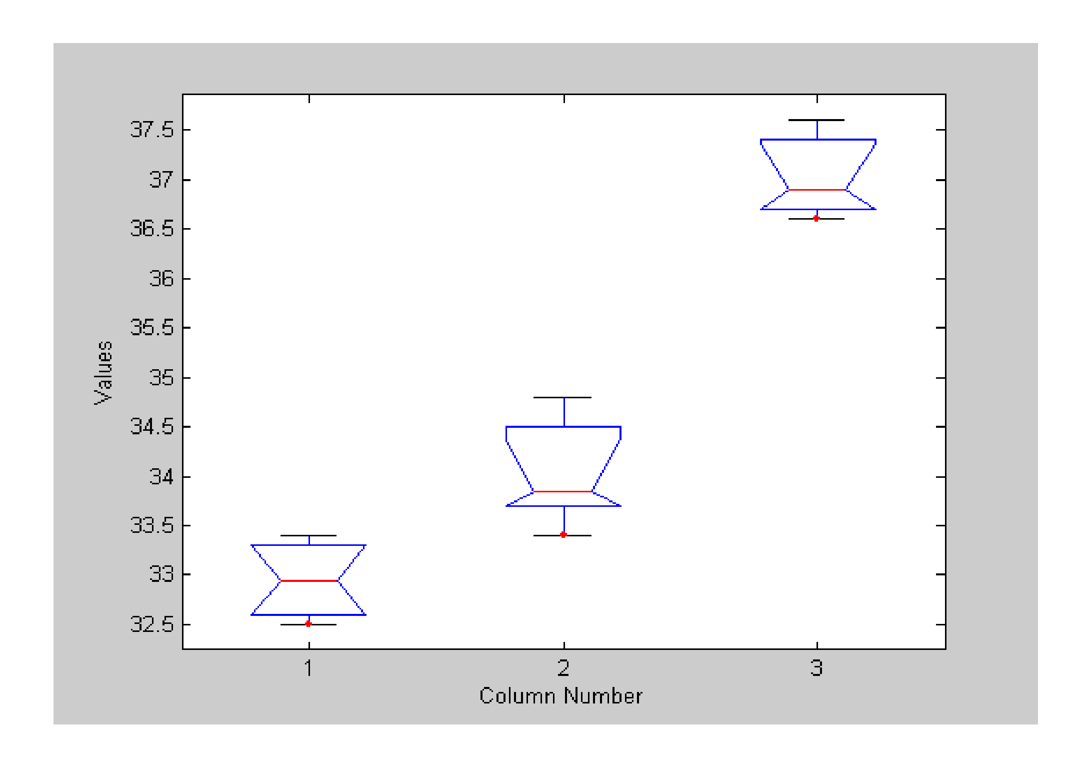

p =

    0.0000          0.0039          0.8411
```

```
tb1 =
'Source'        'SS'          'df'          'MS'      'F'         'Prob>F'
'Columns'            [53.3511]      [  2]      [26.6756]        [234.2244]
[2.4278e-010]
'Rows'          [ 1.4450]       [ 1]     [ 1.4450]       [ 12.6878]    [0.0039]
'Interaction'   [ 0.0400]       [ 2]     [ 0.0200]       [0.1756]      [0.8411]
'Error'              [1.3667]        [12]    [ 0.1139]          []             []
'Total'              [56.2028]       [17]    []                 []             []

stats =
        source: 'anova2'
        sigmasq: 0.1139
        colmeans: [32.9500 34.0167 37.0167]
        coln: 6
        rowmeans: [34.9444 34.3778]
        rown: 9
        inter: 1
        pval: 0.8411
      df: 12
```

المثال السابق وكما موضح في الشكل ٧.٣ يبين أنه توجد ثلاث موديلات للسيارات وهي ممثلة بالأعمدة ويوجد مصنعان ويمثلان بالصفوف. والسبب من وجود ستة صفوف في المثال السابق بدلاً من صفين هو أن كل مصنع تنتج ثلاث سيارات لكل موديل وهي حالة خصصت للدراسة. البيانات من المصنع الأول ممثلة بالصفوف الثلاثة الأولى والبيانات من المصنع الثاني ممثلة بالصفوف الثلاثة الأخيرة.



شكل ٧.٣ مخطط المثال السابق المسافة المقطوعة بالاميال لكل لتر وقود لمصنعان ولثلاث موديلات

يمكن أن تستخدم احصاء F لعمل اختبار الفرضية لايجاد فيما اذا كانت المسافة المقطوعة بالاميال لكل لتر وقود هي نفسها من خلال الموديلات والمصانع وازواج مصنع – موديل حيث أن anova2 ترجع قيمة p من هذه الاختبارات.

حيث أن قيمة p لتأثير الموديل هو صفر وهذا مؤشر قوي لكون المسافة المقطوعة بالاميال لكل لتر من الوقود تتغير من موديل لاخر. أحصاء F هي متطرفة حيث أن القيمة المستحصلة يحب أن تكون أقل من واحد من عشرة آلاف اذا كانت المسافة المقطوعة بالاميال لكل لتر من الوقود هي حقاً متساوية من موديل لموديل آخر.

اذا استخدمت الداله Multcompare لتحقيق اختبار المقارنة المتعدد سوف تجد كل زوج من ثلاث موديلات يختلف بشكل كبير وأن قيمة p بتأثير المصنع هي 0.0039 والتي هي ايضاً مرتفعة وهذا يوضح أن واحد من المصانع يكون خارج مواصفات الاخرين في المسافة المقطوعة بالاميال لكل لتر وقود للسيارات المنتجة. قيمة p تعطي أن احصاء F بكونه غير واقعي من ملاحظة F يحب أن تحصل باربعة لكل 1000 مرة اذا كانت المسافة المقطوعة بالاميال لكل لتر وقود هي في الحقيقة متساوية من مصنع الى مصنع وهذا لا يظهر أي تفاعل بين المصانع والموديلات. قيمة p هي 0.8411 وهذا يعني أن النتيجة الملاحظة هي 84 لكل 100 مرة وهذا لا يعطي أي تفاعل.

قيمة p الراجعة من anova2 تعتمد على الفرضيات عند التوزيعات العشوائية وهي εijk في المعادلة ولتكن قيمة p صحيحة فأن هذه التوزيعات تتطلب أن تكون غير معتمدة وموزعة بشكل طبيعي ولها تباين ثابت.أستخدام الدالة anova2 يتطلب أن تكون البيانات متوازنة وهذا يعني تستخدم نفس العدد من السيارات لاي تشكيله من الموديلات والمصانع.

٧.٦ تحليل التباين متعدد الطرق N-way Analysis of Variance

تحليل التباين متعدد الطرق يمكن أستخدامه لحساب إذا كانت المعدلات لمجموعة من البيانات تختلف عن بعضها وذلك بتقسمها حسب عوامل متعددة. وعند ذلك يمكنك أن تحسب أي العوامل أو تشكيلة العوامل لها علاقة بالاختلاف. تحليل التباين متعدد الطرق هو تعميم لتحليل التباين ذو طريقين فالمعادلة لثلاث عوامل يمكن أن تكون كما يلي:

$$yijkl = \mu + \alpha j + \beta i + \gamma k + (\alpha\beta)ij + (\alpha\gamma)ik + (\beta\gamma)jk + (\alpha\beta\gamma)ijk + \varepsilon ijkl$$

ففي هذه المعادلة فان العناصر التي تحتوي على متغيرين $(\alpha\beta)ij$ تمثل تأثير التفاعلات لعاملين وأن العناصر التي تحتوي على ثلاث متغيرات $(\alpha\beta\gamma)ijk$ تمثل تأثير التفاعلات لثلاث عوامل. وبأستخدام نموذج anova يمكن أن يكون له مجموعة كاملة من المتغيرات أو أي مجاميع ولكنه لايشمل تفاعلات معقدة الا اذا كان يحتوي على تفاعلات ذات ثلاث طرق.

فلو تم توظيف هذه العلاقة في تحليل التباين متعدد الطرق وادخال كافة العوامل المؤثرة بالعلاقة فبالامكان تكوين خوارزمية أو نظام معلومات يضم كافة العناصر وهذا يؤدي الى تحقيق نتائج أفضل من حيث دراسة تأثير كافة التفاعلات ولكافة العوامل وهذا يدخل بشكل كبير ويساعد على أتخاذ القرار الصحيح من خلال أدخال كافة المؤثرات في النظام المعلوماتي.

٧.٧ تطبيق تحليل التباين متعدد الطرق لمجموعة صغيرة

Example of small group N-way Analysis of Variance

يمكن تطبيق تحليل التباين لمجموعة صغيرة من البيانات وكمثال على ذلك نأخذ مصفوفة ٣*٤ وكما يلي :

```
>> m = [ 23   15   20  ;  27   17   63  ;  43   3   55  ;  41   9   90 ]


m=

        23      15      20

        27      17      63

        43       3      55

        41       9      90
```

نطبق الدالة anova2 على المصفوفة m وكما يلي:

```
>> anova2 (m, 2)


ans = 0.0197    0.22234    0.2663
```

المعلومات التي تدل على العامل تطبق على شكل المصفوفة m وعدد القياسات لكل تشكيلة بالنسبة الى العامل وبالرغم أن anova2 لا تتطلب في الحقيقة صفوف بقيم العوامل ولتوضيح ذلك يمكن أن نتبع الخطوات التالية:

أولا نولد العوامل باستخدام الدالة repmat وكما يلي:

```
>>  cfactor = repmat (1:3, 4, 1)

 cfactor =    1    2    3
              1    2    3
              1    2    3
              1    2    3
```

الدالة بالقيم المعطاة تولد مصفوفة بعدد الاعمدة ثلاثة وعدد الصفوف أربعة وعدد التكرارات واحد حيث يكون العمود الأول عبارة عن واحد والعمود الثاني عبارة عن اثنين والعمود الثالث عبارة عن ثلاثة.

نولد مصفوفة أخرى بقيم أخرى بأستخدام الدالة التالية:

```
rfactor = [ones (2, 3); 2*ones (2, 3)]

rfactor =    1    1    1
             1    1    1
             2    2    2
             2    2    2
```

الدالة بالقيم المعطاة تولد مصفوفة يكون الجزء الاول فيها قيم واحد لصفين وثلاثة أعمدة أما الجزء الثاني فيها فقد تم ضرب نتيجة الجزء الاول في أثنين. حيث يكون الصفين العلويين من المصفوفة يحمل القيمة واحد والصفين السفليين من المصفوفة يحمل القيمة اثنان، وبتعبير آخر كل قيمة للمصفوفة m(i, j) تمثل مستوى عامل الأعمدة cfactor(i, j) ومستوى عامل الصفوف rfactor(i,j) ولحل هذه المسألة باستخدام الدالة anovan يجب أن تعيد هيكلة المصفوفات m و cfactor و rfactor لتكون متجهات وكما يلي:

```
>> m = m(:)          cfactor = cfactor(:)          rfactor = rfactor(:)

rfactor =            cfactor =                     m =

١                    ١                             ٢٣
١                    ١                             ٢٧
٢                    ١                        ٤٣
٢                    ١                             ٤١
١                    ٢                             ١٥
١                    ٢                             ١٧
٢                    ٢                             ٣
٢                    ٢                             ٩
١                    ٣                             ٢٠
١                    ٣                             ٦٣
٢                    ٣                             ٥٥
٢                    ٣                             ٩٠
```

ثم نعيد تجميع المصفوفات الثلاثة في مصفوفة واحدة وكما يلي:

```
>> [ m    cfactor    rfactor ]

= ans

 ١   ١   ٢٣
 ١   ١   ٢٧
 ٢   ١   ٤٣
 ٢   ١   ٤١
 ١   ٢   ١٥
```

```
١   ٢   ١٧
٢   ٢   ٣
٢   ٢   ٩
١   ٣   ٢٠
١   ٣   ٦٣
٢   ٣   ٥٥
٢   ٣   ٩٠
```

بعد ذلك نطبق الدالة anovan وكما يلي:

```
>> anovan (m, {cfactor  rfactor}, 2)


ans = 0.0197
      0.2234
      0.2663
```

٧.٨ تطبيق تحليل التباين متعدد الطرق لمجموعة كبيرة

Example of large group N-way Analysis of Variance

لتطبيق تحليل التباين متعدد الطرق لمجموعة كبيرة نستعرض المثال التالي الذي يوضح كيفية تحليل مجموعة كبيرة من بيانات السيارات مقارنة المسافة المقطوعة بالأميال لكل لتر وقود ومعلومات أخرى لسيارات نوع 406 المصنوعة بين عامي 1970 و 1982. كبداية نحمل البيانات من الملف carbig باستخدام الدالة load ثم نجد ماذا تحتوي الأعمدة وكما يلي:

```
>> load carbig
>> whos
```

Name	Size	Bytes	Class
Acceleration	406x1	3248	double array
Cylinders	406x1	3248	double array
Displacement	406x1	3248	double array
Horsepower	406x1	3248	double array
MPG	406x1	3248	double array
Model	406x36	29232	char array
Model_Year	406x1	3248	double array
Origin	406x7	5684	char array
Weight	406x1	3248	double array
Ans	3x1	24	double array
c	12x1	96	double array
cyl4	406x5	4060	char array
m	12x1	96	double array
org	406x7	5684	char array
r	12x1	96	double array
when	406x5	4060	char array

المثال يركز على أربع متغيرات هي:

MPG ويعني عدد الأميال المقطوعة لكل غالون لكل سيارات 406 .

cyl4 ويعني كون السيارات بأربع أسطوانات أم لا.

org ويعني مواصفات السيارات على كونها أوربية أو يابانية أو أمريكية.

when ويعني السيارة صنعت في الفترة الزمنية أو في منتصفها أو في نهايتها.

نطبق النموذج الشامل مع التفاعلات ثلاثية الطرق والنتائج تظهر في الشكل ٧.٤ وتنفذ الاوامركما يلي:

```
>> varnames = {'origin'; '4 cyl'; 'MfgDate'};
>> anovan (MPG, {org  cyl4  when}, 3, 3, varnames)
    ans =
          0.0000
          Non
          0
          0.7032
          0.0001
          0.2072
          0.6990
```

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
# origin	416.8	1	416.77	29.34	0
# 4 cyl	0	0	0	0	NaN
# MfgDate	1112.3	1	1112.27	78.31	0
# origin*4 cyl	2.1	1	2.07	0.15	0.7032
# origin*MfgDate	301.2	3	100.41	7.07	0.0001
# 4 cyl*MfgDate	22.7	1	22.68	1.6	0.2072
# origin*4 cyl*MfgDate	20.3	3	6.77	0.48	0.699
Error	5411.8	381	14.2		
Total	24252.6	397			

Constrained (Type III) sums of squares. Terms marked with # are not full rank.

شكل ٧.٤ جدول التباين بتطبق النموذج الشامل مع التفاعلات ثلاثية الطرق

من تقاطع الجداول أعلاه يبين انه لا توجد سيارات مصنوعة في اوربا في الفترة المبكرة من الزمن المحدد غير النوع ذو أربع اسطوانات والتـي تكـون مـؤشرة بالقيمـة صـفر ففي البداية ننشأ الجدول باستخدام دالة التقاطع crosstab وكما يلي:

```
>> [table, chi2, P, factorvals] = crosstab (org, when, cyl4)

table ( :, :, 1) =

        82    75    25

         0     4     3

         3     3     4

table ( :, :, 2) =

        12    22    38

        23    26    17

        12    25    32

chi2 =

        207.7689

p =

   0

factorvals =

      'USA'      'Gorly'     'Other'

      'Europe'   'Mid'       'four'

      'Japan'    'Late'      [ ]
```

باستخدام المعلومات المتوفرة والمحدودة في جدول anova يمكن أن تلاحظ أن التفاعل ثلاثي الطرق له قيمة P تساوي 0.699 وهي ليست ذات قيمة لذلك يمكنك أن تقوم باحتساب فقط تفاعلات ثنائية الطرق.

نطبق دالة anovan حيث نحصل على النتيجة في الشكل ٧.٥ وتنفذ كما يلي:

```
>> [p, tbl, stats, termvec]= anovan( MPG, {org  cyl4  when}, 2,3,
          varname);

p =
٠.٠٠٠٠
٠
٠
٠.٦٤٢٢
٠.٠٠٠١
٠.٣٣٤٨

= tbl
Columns 1 through 3
'.Source'            'Sum Sq.'          'd.f'
'origin'             [   532.5813]       [ 2  ]
'cyl'                [1.7698e+003]      [ 1٤ ]
'MfgDate'            [2.8871e+003]      [ 2  ]
'origin*4 cyl'       [12.5435]          [ 2 ]
'origin*MfgDate'     [350.3555]         [ 4  ]
'cyl*MfgDate'        [  31.0477]        [2 ٤ ]
```

'Error'	[5.4321e+003]	[384]
'Total'	[2.4253e+004]	[397]

Columns 4 through 6

'Singular?'	'Mean Sq.'	'F'
[١٨.٨٢٤٤]	[٢٦٦.٢٩٠٧]	[٠]
[e+003]	[125.11231.٧٦٩٨]	[٠]
[e+003]	[102.04581.٤٤٣٥]	[٠]
[٠.٤٤٣٤]	[٦.٢٧١٨]	[٠]
[٦.١٩١٨]	[٨٧.٥٨٨٩]	[٠]
[١.٠٩٧٤]	[١٥.٥٢٣٨]	[٠]
[]	[١٤.١٤٦١]	[٠]
[]	[]	[٠]

Column 7

'Prob>F'

[e-0081.٥٨٨٧]

[٠]

[٠]

[٠.٦٤٢٢]

[e-005٧.٧٥٨٠]

[٠.٣٣٤٨]

```
[]
[]

= stats
'source: 'anovan
[coeffs: [14x1 double
[Rtr: [14x14 double
dfe: 384
mse: 14.1461
[permvec: [1 4 5 13 7 14 6 3 10 11 2 9 12 8
[terms: [6x1 double
[nlevels: [3x1 double
[termcols: [7x1 double
= termvec
        ١
        ٢
        ٤
        ٣
        ٥
        ٦
```

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
origin	532.6	2	266.29	18.82	0
4 cyl	1769.8	1	1769.85	125.11	0
MfgDate	2887.1	2	1443.55	102.05	0
origin*4 cyl	12.5	2	6.27	0.44	0.6422
origin*MfgDate	350.4	4	87.59	6.19	0.0001
4 cyl*MfgDate	31	2	15.52	1.1	0.3348
Error	5432.1	384	14.15		
Total	24252.6	397			

Constrained (Type III) sums of squares.

شكل ٧.٥ جدول التباين للمثال السابق متعدد الطرق

الآن كل العناصر تكون قد تم خمنينها حيث أن قيمة P للتفاعل 4 (origin * 4 cyl) وللتفاعل 6 هي (4 cyl * MfgDate) وهي أكبر بكثير من قيمة القطع المددة 0.05 علما بأن هذه القيم ليست فعالة وأن ناتج المتغير termvec يرجع الصف إلى تمثيل قيمة البت.

بامكانك حذف حدود من النموذج بحذفهم من termvec وتنفذ anovan من جديد وفي هذه الحالة يجهز صف النتيجة كما في الشكل ٧.٦ وتنفذ الاوامركما يلي:

```
>> termvec ([ 4   6 ], :) = [ ]

termvec =

        ١
        ٢
        ٤
        ٥
```

```
>> anovan (MPG, {org   cyl4   when}, termvec, 3,varnames)

Ans =

    1.0e – 003

    0.0000

        0

        0

     0.1140
```

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
origin	686.7	2	343.36	24.34	0
4 cyl	4206.2	1	4206.17	298.19	0
MfgDate	3590.7	2	1795.34	127.28	0
origin*MfgDate	336.8	4	84.19	5.97	0.0001
Error	5473	388	14.11		
Total	24252.6	397			

Constrained (Type III) sums of squares.

شكل ٧.٦ جدول التباين المتعدد في المثال السابق بأستخدام anovan

٧.9 تحليل التباين بالتأثيرات العشوائية ANOVA with Random Effects

يتميز نموذج anova ألاعتيادي بكون كل مجموعة متغيرة تمثل عامل ثابت وان مستويات هذا العامل مجموعة ثابتة للقيم. الهدف من التطبيق هو حساب فيما إذا كانت مستويات العامل المختلفة تقود إلى قيم للاستجابة المختلفة. وسوف نستعرض كيفية استخدام anovan لتقريب النماذج حيث أن مستويات العامل يمثل أختيار عشوائي من مجموعة كبيرة لمستويات الاحتمال. وهنا سوف نستعرض عدد من الفقرات كما يلي:

- تجهيز النموذج.
- توفيق نموذج التأثيرات العشوائية.
- إحصاء F للنماذج بالتأثيرات العشوائية.
- مركبات التباين.

فلو تم توطين هذه الفقرات من تحليل التباين بالتأثيرات العشوائية في خوارزمية تضم كافة الخطوات والتأثيرات لأصبح هناك نظام معلوماتي مميز وبإمكانه حل كافة النماذج ذات العلاقة.

٧.٩.١ تجهيز النموذج Setting Up the Model

تتضمن تجهيز النموذج كافة الامور المتعلقة بتهية النموذج وهي تحميل البيانات باستخدام الدالة load ولتحميل الملف mileage على سبيل المثال وهو عبارة عن مصفوفة 3 x 6 تمثل المسافة المقطوعة بالأميال نكتب الامر وكما يلي:

```
>> load mileage
```

الدالة anova2 تعمل فقط مع البيانات المتوازنة وهي تدل على القيم لمتغيرات المجموعة من إعداد الصفوف والأعمدة لمصفوفة الإدخال. الدالة anovan تتطلب إنشاء أو تكوين صفوف من قيم من متغيرة المجموعة ولتكوين هذه الصفوف تتبع الخطوات التالية:

١- إيجاد وتكوين صف يدل على عامل لكل قيمة في الملف وهـذا الصـف هـو واحـد للعمود الأول واثنان للعمود الثاني وثلاثة للعمود الثالث وهكذا وكما يلي:

```
>> factory = repmat (1 : 3, 6, 1)

factory =


    1    2    3

    1    2    3

    1    2    3

    1    2    3

    1    2    3

    1    2    3
```

٢- إيجاد وتكوين صف يدل على نموذج السيارة لكـل قيمـة مـن المسـافة المقطوعـة وهذا الصف هو واحد للصف الأول من الملف واثنان يمثل الصفوف المتبقية .

```
>> carmod = [ ones (3,3); 2 * ones (3,3) ]


= carmod


    ١    ١    ١

    ١    ١    ١

    ١    ١    ١

    ٢    ٢    ٢

    ٢    ٢    ٢

    ٢    ٢    ٢
```

٣- تحويل هذه المصفوفات إلى متجهات وعرضها كما يلي:

```
>> mileage = mileage ( : ) ;

>> factory = factory  ( : ) ;

>> carmod = carmod ( : ) ;

>> [mileage  factory  carmod]

= ans

١.٠٠٠٠   ١.٠٠٠٠   ٣٣.٣٠٠٠
١.٠٠٠٠   ١.٠٠٠٠   ٣٣.٤٠٠٠
١.٠٠٠٠   ١.٠٠٠٠   ٣٢.٩٠٠٠
٢.٠٠٠٠   ١.٠٠٠٠   ٣٢.٦٠٠٠
٢.٠٠٠٠   ١.٠٠٠٠   ٣٢.٥٠٠٠
٢.٠٠٠٠   ١.٠٠٠٠   ٣٣.٠٠٠٠
١.٠٠٠٠   ٢.٠٠٠٠   ٣٤.٥٠٠٠
١.٠٠٠٠   ٢.٠٠٠٠   ٣٤.٨٠٠٠
١.٠٠٠٠   ٢.٠٠٠٠   ٣٣.٨٠٠٠
٢.٠٠٠٠   ٢.٠٠٠٠   ٣٣.٤٠٠٠
٢.٠٠٠٠   ٢.٠٠٠٠   ٣٣.٧٠٠٠
٢.٠٠٠٠   ٢.٠٠٠٠   ٣٣.٩٠٠٠
١.٠٠٠٠   ٣.٠٠٠٠   ٣٧.٤٠٠٠
```

١.٠٠٠٠	٣.٠٠٠٠	٣٦.٨٠٠٠
١.٠٠٠٠	٣.٠٠٠٠	٣٧.٦٠٠٠
٢.٠٠٠٠	٣.٠٠٠٠	٣٦.٦٠٠٠
٢.٠٠٠٠	٣.٠٠٠٠	٣٧.٠٠٠٠
٢.٠٠٠٠	٣.٠٠٠٠	٣٦.٧٠٠٠

٧.٩.٢ موائمة نموذج التأثيرات العشوائية **Fitting a Random Effects Model**

لو فرضنا انك أعددت دراسة على عدد من المصانع حيث أنك ترغب بان المعلومات حول ما يحدث تبنى على نفس موديلات السيارات في مصانع مختلفة. للحصول على هذه المعلومات يتم موائمة وموافقة تحليل نموذج التباين ثم بعدها نحدد النموذج الذي يحتوي على حد التفاعل والذي يكون عشوائي بالنسبة الى عامل المصنع وكما موضحة في الشكل ٧.٧ وتكون الاوامركما يلي :

```
>> [pvals, tbl, stats] = anovan(mileage, {factory carmod}, ...

'model',2, 'random',1,'varnames',{'Factory' 'Car Model'});
```

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Factory	53.3511	2	26.6756	1333.78	0.0007
Car Model	1.445	1	1.445	72.25	0.0136
Factory*Car Model	0.04	2	0.02	0.18	0.8411
Error	1.3667	12	0.1139		
Total	56.2028	17			

Constrained (Type III) sums of squares.

شكل ٧.٧ جدول التباين للمثال السابق بأستخدام الدالة anovan

٧.٩.٣ أحصاء F للنماذج بالتأثيرات العشوائية

F statistic for Models with Random Effects

نموذج أحصاء F له تأثيرات عشوائية ويعرف بشكل مختلف عن نموذج يحتوي كل التأثيرات الثابتة. يحسب أحصاء F في نموذج التأثيرات الثابتة لاي حد بأخذ النسبة لمعدل التربيع لذلك الحد بمعدل تربيع للخطأ علما بأنه في نموذج التأثيرات العشوائية فأن بعض أحصاءات F يستخدم معدل تربيع في المقام. المثال الموضح في نموذج التحميل يبين أن تأثير المتغير 'Factory' يجب أن يتغير حول موديلات السيارات، وفي هذه الحالة فأن تفاعل معدل التربيع يحدث بالنسبة الى خطأ معدل التربيع في أحصاء F، وأن أحصاء F للمصنع هو:

F = 1.445/0.02

F= 72.2500

درجات الحرية للاحصاء هي درجات الحرية للبسط (١) والمقام (٢) معدل المربعات وكما يلي :

```
>> Pval = 1 – fcdf (F , 1 , 2)


Pval = 0.0136
```

بتطبيق التأثيرات العشوائية فان القيمة المتوقعة لكل مربع المعدل تعتمد ليس فقط على التباين لحد الخطأ ولكن كذلك للتباينات الموزعة بالتأثيرات العشوائية. يمكن ان تعرض هذه الاعتمادات بكتابة القيم المتوقعة بتشكيلات خطية للاستنتاجات من حدود النموذج المختلفة لايجاد العوامل لهذه التشكيلات الخطية أولا ندخلً stats.ems الذي يرجع حقل ems لهيكل stats وكما يلي:

```
>> stats.ems

ans =

    1.0000    6.0000    0.0000    3.0000
    1.0000    0.0000    9.0000    3.0000
    1.0000    0.0000    0.0000    3.0000
    1.0000         0         0         0
```

لايجاد تمثيل المتن للتشكيلات الخطية فندخلها كما يلي :

```
>> stats.txtems

ans =

'6 * V (Factory) + 3 * V (Factory * car Models) + (Error)'

'9 * Q (car Model) + 3 * V (Factory * car Model) + V (Error)'
'3 * V (Factory * car Model) + V ( Error)
 'V (Error)'
```

القيمة المتوقعة لمربع المعدل خلال موديل السيارة (الحد الثاني) ويتضمن استنتاجات من دالة رباعية لتأثيرات موديل السيارة زائداً ثلاثة اضعاف التباين لتأثير حد التفاعل زاداً التباين لحد الخطأ اخذين بنظر الاعتبار اذا كان تأثيرات موديل السيارة يكون صفراً فان الحد يقل الى القيمة المتوقعة لمعدل المربع للحد الثالث (حد التفاعل) ، وهذا يعطي سبب كون احصاء F لتأثير موديل السيارة يستخدم تفاعل معدل المربع في المقام.

في بعض الحالات لا يوجد حد واحد تكون قيمته المتوقعة توافق القيمة المطلوبة للمقام لاحصاء F ، وفي تلك الحالة المقام يكون تشكيلة خطية لمربع المعدلأت. هيكله stats يحتوي على حقول يعطي تعريفات للمقامات لكل أحصاء F. حقل txtdenom و stats. txtdenom يمثلان المتن وحقل denom يعطي المصفوفه التي تعرف تشكيلة خطية لتباينات الحدود في النموذج. للنماذج المتوازنة كالذي عرض فان مصفوفة demon و stats. denom يحتويان أصفار وواحدات بسبب ان المقام هو حد مفرد لمربع المعدل وكما يلي :

```
>> stats . txtdenom

ans =

'MS (Factory * car Model)'

'MS (Factory * car Model)'

'MS (Error)'

>> stats. denom

ans =
0.0000      1.0000      0.0000
0.0000      1.0000      0.0000
  0           1.0000      0.0000
```

٧.٩.٤ مركبات التباين Variance Components

في النموذج الموضح في نموذج التحميل نأخذ بنظر الاعتبار المسافة المقطوعة للسيارات الخاصة من الموديلات الخاصة المصنوعة في معمل عشوائي، حيث أن التباين في تلك السيارات هو مجموعة المركبات او الاستنتاجات واحد من الحدود العشوائية وكما يلي :

```
>> stats . rtnames

ans =

'Factory'
'Factory * car Model'
'Error'
```

هنا لا تعرف تلك التباينات ولكن يمكنك تخمينها من البيانات وهذا يدعى حقل ems لهيكل stats ويوضح القيمة المتوقعة لكل حد من مربع المعدل لتشكيله خطية للتباينات الغير معروفة للحدود العشوائية والهيئات الرباعية الغير معروفة للحدود الثابتة. لو أخذت حدود مربع المعدل المتوقعة وساويتها مع تلك القيم المتوقعة لمربع المعدلات المحسوبة فبأمكانك ان تحصل على نظام للمعادلات التي يمكن حلها للتباينات الغير معروفة وهذه الحلول هي تخمينات لمركبة التباين والحقل varest يحتوي على اسماء للحدود العشوائية وكما يلي :

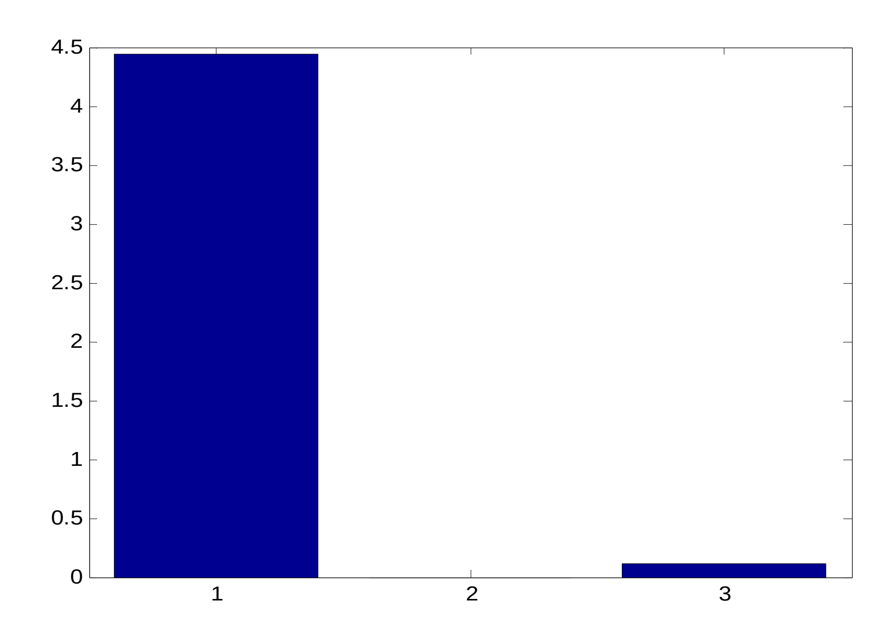
```
>> stats.varest
ans =
 4.4426
        -0.0313
         0.1139
```

تحت بعض الشروط فان التغير الى الحد يكون غالباً قليل وتخمن مركبة التباين بانها سالبة ومن المألوف في هذه الحالات وضع التخمين لصفر والذي تعمله لتكوين مخطط المستطيلات للمركبات كما في شكل ٧.٨ وتكتب كما يلي:

```
>> bar (max (0, stats.varest))

set (gca, 'xtick' , 1 : 3, 'xticklabel', stats.rtnames)

bar (max (0, stats . varest))
```



شكل ٧.٨ مخطط المستطيلات للمثال السابق

يمكنك أيضاً حساب نقاط الثقة لتخمين التباين، حيث أن الدالة anovan بأمكانها ان تعمل ذلك بحساب نقاط الثقة للتباين المتوقع لمربع المعدلات وايجاد النهاية الصغرى والنهاية العظمى لكل مركبة من مركبات التباين التي تحتوي على جميع هذه النقاط. الطريقة هذه تقود الى مجموعة من النقاط المستخلصة للبيانات المتوازنة وهذا يعني ان 95% من نقاط الثقة لها احتمالية على الاقل 95% من محتويات التباينات الحقيقية اذا كان عدد الاستنتاجات لكل تشكيلة من متغيرات المجموعة هي نفسها. للبيانات الغير متوازنة هذه هي تقريبات وهي غير مضمونة لتكون استنتاجات وهي كما يلي :

```
>> [ {'Term' 'Estimate' 'Lower' 'Upper'};

        stats.rtname, num2cell ( [stats.varest stats.varci] ) ]
```

'Term'	'Estimate'	'Lower'	'Upper'
'Factory'	[4.4426]	[1.0736]	[175.6038]
'Factory + car Model'	[-0.0313]	[NaN]	[NaN]
'Error'	[0.1139]	[0.0586]	[0.3103]

الفصل الثامن

تطبيقات حزمة MATLAB

في النماذج الخطية / التحليلات الاخرى

Matlab Application on

Linear Models / Other Analysis

٨.١ مقدمة Introduction

التحليلات الاخرى للنماذج الخطية في حزمة MATLAB هي جميع التحليلات التي لم يرد ذكرها في الفصل السابق أي أنها عدا تحليل التباين الذي تم استعراضه في سابقا وفي هذا الفصل سوف نستعرض مجموعة من الفقرات وتشمل ما يلي:

١. تحليل التباين المشترك Analysis of Covariance

٢. الانحدارالخطي المتعدد Multiple Linear Regression

٣. نماذج سطح الاستجابة التربيعي Quadratic Response Surface Models

٤. الانحدار ذو الخطوات المتعاقبة Stepwise Regression

٥. النماذج الخطية العامة Generalized Linear Models

٦. الطرق اللامعلمية النشيطة Robust and Nonparametric Methods

٨.٢ تحليل التغايير Analysis of Covariance

تحليل التغايير covaniance هو تقنية لتحليل البيانات المجمعة لها استجابة نرمز لها بالمتغير التابع y (وهو المتغير الذي يتم التنبأ به) ولها متغير مستقل نرمز له بالمتغير x (وهو المتغير المستخدم لعمل التنبأ). بأستخدام تحليل conviniace يمكن نمذجة y كدالة خطية للمتغير x.

يوجد في حزمة MATLAB برنامج جاهز هو الدالة aoctool والذي يمثل تفاعل بين المتغيرات المستقلة وبأستخدام تحليل نماذج covariance نطبق الدالة anocava وهو مشابه الى البرنامج الجاهز polytool. وان الدالة aoctool توافق النماذج التالية الى المجموعة:

- نفس الوسط الحسابي same mean
- عزل الاوساط الحسابية separate means
- الخط نفسه same line

- الخطوط المتوازية parallel lines
- الخطوط المعزولة separate line

ففي نموذج الخطوط المتوازية مثلا ان الاعتراض والتوقف يتغير من مجموعة الى اخرى لكن الميل هو نفسه لكل مجموعة ولكي تكون عوامل المجموعة محسوبة بشكل صحيح فأن البرنامج الجاهز يفرض المتغيرات.

ولتطبيق الدالة aoctool نتبع الخطوات التالية:

١. تحميل البيانات وذلك من خلال الحزمة الاحصائية حيث يوجد الملف carsmall والذي يمثل بيانات عن السيارات للسنوات ١٩٧٠، ١٩٧٦، ١٩٨٢ وهذا المثال يدرس العلاقة بين أوزان السيارات والمسافة التي تقطعها وفيما اذا كانت هذه العلاقة تغيرت عبر السنوات وكبداية نحمل الملف carsmall وكما يلي:

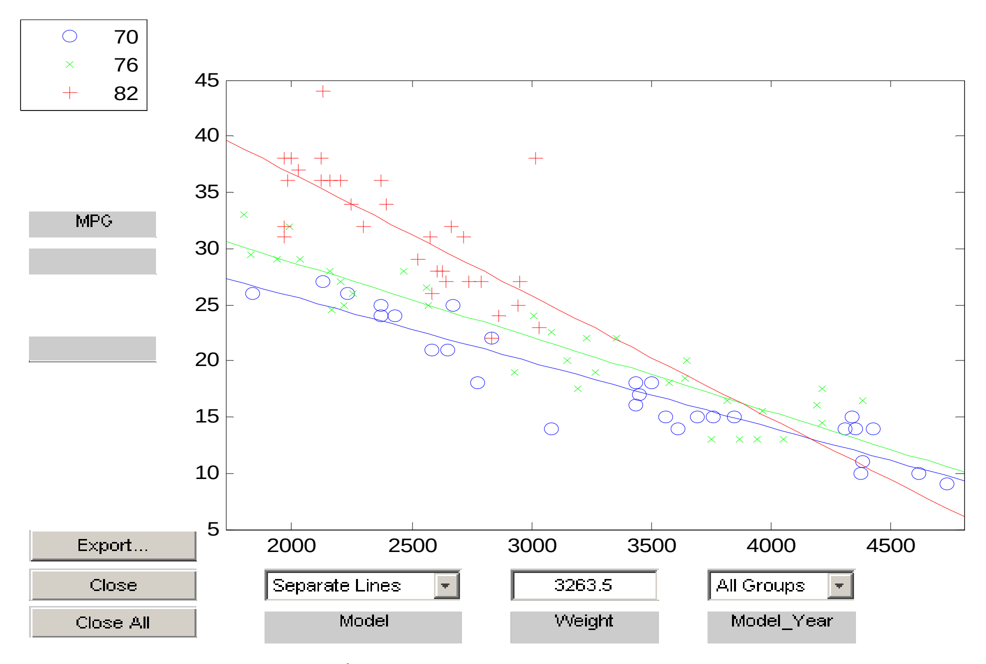
```
>> load carsmall
```

وكذلك بأمكانك استخدام الدالة aoctool بالبيانات التي تحددها.

٢. بدأ استخدام الأداة حيث أن الأوامر التالية للدالة aoctool لتوفيق الخط المنفصل لمتجهات الاعمدة الوزن (weight) والمسافة المقطوعة بالاميال لكل غالون (MPG) واحد من مجموعة الموديلات الثلاثة والمعرفة Model-Year . أن توافق النماذج الاساسية هي المتغير MPG، y كدالة خطية للمتغير x والوزن weight وكما يلي:

```
>> [h,atab,ctab,stats] = aoctool(Weight,MPG,Model_Year);
Note: 6 observations with missing values have been removed
```

٣. اختبر الإخراج حيث انه يحتوي على نافذة رئيسية والتي تمثل بالرسم البياني شكل ٩.١ ففي الرسم كل مجموعة Model-Year لها خط منفصل ونقاط البيانات لكل مجموعة لها نفس اللون كنقاط للبيانات أما الاخراج الثاني فهو جدول للعوامل التخمينية شكل ٩.٢ وأما الأخراج الثالث فهوتحليل لجدول التباين شكل ٨.١.

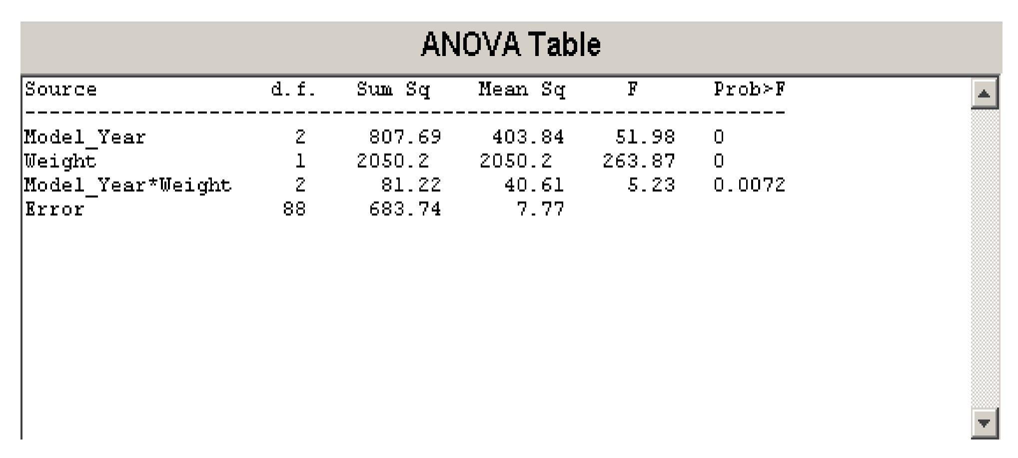


شكل ٨.١ الرسم البياني لملف السيارات carsmall بأستخدام الدالة aoctool

Coefficient Estimates

Term	Estimate	Std. Err.	T	Prob>|T|
Intercept	45.9798	1.52085	30.23	0
70	-8.5805	1.96186	-4.37	0
76	-3.8902	1.86864	-2.08	0.0403
82	12.4707	2.5568	4.88	0
Slope	-0.0078	0.00056	-14	0
70	0.002	0.00066	2.96	0.0039
76	0.0011	0.00065	1.74	0.0849
82	-0.0031	0.001	-3.1	0.0026

شكل ٨.٢ جدول العوامل لملف السيارات carsmall بأستخدام الدالة aoctool



ANOVA Table

Source	d.f.	Sum Sq	Mean Sq	F	Prob>F
Model_Year	2	807.69	403.84	51.98	0
Weight	1	2050.2	2050.2	263.87	0
Model_Year*Weight	2	81.22	40.61	5.23	0.0072
Error	88	683.74	7.77		

شكل ٨.٢ جدول أختبار النتائج لملف السيارات carsmall بأستخدام الدالة aoctool

عوامل الخطوط الثلاثة تظهر في الشكل المعنون عوامل ANOCOVA وان الميل للكل يكون بشكل تقريبي ٠.٠٠٧٨ - مع انحراف قليل لكل مجموعة للموديل ١٩٧٠ والموديل ١٩٧٦ والموديل ١٩٨٢.

٨.٣ الانحدار الخطي المتعدد Multiple Liner Regression

الهدف من الانحدار الخطي المتعدد هو لايجاد علاقة نوعية بين مجموعة المتغيرات المستقلة وهي أعمدة x والاستجابة للمتغير التابع y وهذه العلاقة مفيدة وكالاتي:

- فهم أي من المتغيرات المستقلة لها تأثير اكبر.
- معرفة اتجاه التأثير.
- استخدام النموذج للتبأ بالقيم المستقبلية للاستجابة عندما تكون فقط المتغيرات المستقلة معروفة.

والفقرات التالية توضح بشكل تفصيلي الانحدار الخطي المتعدد.

٨.٣.١ البناء الرياضي لنموذج الانحدار الخطي المتعدد

Mathematical Foundations of Multiple Liner Regression

النموذج الخطي يأخذ الهيئة المشتركة كما في المعادلة:

$$y = X\beta + \varepsilon$$

حيث ان

y: تمثل متجه n*1

X: مصفوفة الانحدار n*p

β: تمثل متجه p*1 للمعلمات

ε: تمثل متجه n*1 للتوزيعات العشوائية (الخطأ)

حل هذه المشكلة هو متجه b والذي يمثل القيم التقديرية لمتجه المعلمات β وبأستخدام طريقة المربعات الصغرى الذي يتمثل في المعادلة التالية:

$$b = (X^TX)^{-1} X^Ty$$

هذه المعادلة هي مفيدة لتطوير المعادلات الاحصائية الأخرى ولكن لها خواص رقمية فقيرة.

٨.٣.٢ تطبيق على الانحدار الخطي المتعدد

Example of multiple linear regression

لتطبيق الانحدار الخطي المتعدد نحمل الملف Moore وهو موجود ضمن البرنامج الاحصائى في حزمة MATLAB وذلك بأستخدام الدالة load وبعدها نكون المصفوفة X وتكون الاوامركما يلي:

```
>> lood moore
>> X = [ones(size(moore,1),1) moore(:,1:5)]
```

X =

1.0e+003 *

0.0010	1.1250	0.2320	7.1600	0.0859	8.9050
0.0010	0.9200	0.2680	8.8040	0.0865	7.3880
0.0010	0.8350	0.2710	8.1080	0.0852	5.3480
0.0010	1.0000	0.2370	6.3700	0.0838	8.0560
0.0010	1.1500	0.1920	6.4410	0.0821	6.9600
0.0010	0.9900	0.2020	5.1540	0.0792	5.6900
0.0010	0.8400	0.1840	5.8960	0.0812	6.9320
0.0010	0.6500	0.2000	5.3360	0.0806	5.4000
0.0010	0.6400	0.1800	5.0410	0.0784	3.1770
0.0010	0.5830	0.1650	5.0120	0.0793	4.4610
0.0010	0.5700	0.1510	4.8250	0.0787	3.9010
0.0010	0.5700	0.1710	4.3910	0.0780	5.0020
0.0010	0.5100	0.2430	4.3200	0.0723	4.6650
0.0010	0.5550	0.1470	3.7090	0.0749	4.6420
0.0010	0.4600	0.2860	3.9690	0.0744	4.8400
0.0010	0.2750	0.1980	3.5580	0.0725	4.4790
0.0010	0.5100	0.1960	4.3610	0.0577	4.2000
0.0010	0.1650	0.2100	3.3010	0.0718	3.4100
0.0010	0.2440	0.3270	2.9640	0.0725	3.3600
0.0010	0.0790	0.3340	2.7770	0.0719	2.5990

المصفوفة X كل حدودها مضروبة في اللرقم ١٠٠٠ للتسيل الحسابي لها عمود من الواحدات وعمود آخر من القيم لكل المتغيرات الخمس وان عمود الواحدات هو ضروري لتقدير توافق y للنموذج الخطي وكما يلي :

```
>> y = moore(:,6)

y =

   1.5563    0.8976    0.7482    0.7160    0.3130
   0.3617    0.1139    0.1139   -0.2218   -0.1549
 -0.0969   -0.2218   -0.3979    0              0
  -0.1549   -0.2218   -0.3979   -0.5229   -0.0458
```

والان نستخدم الدالة regress وتعطي الانحدار الخطي المتعدد قيم المربعات الصغرى وكما يلي:

```
>> [b,bint,r,rint,stats] = regress(y,X)

b =

  -2.1561   -0.0000    0.0013    0.0001    0.0079    0.0001

bint =

  -4.1154   -0.1969
  -0.0011    0.0011
  -0.0014    0.0040
  -0.0000    0.0003
  -0.0221    0.0379
  -0.0000    0.0003
```

```
r =

    0.5623   -0.1456    0.0885   -0.0479   -0.2307
    0.1707   -0.3413   -0.0708   -0.0103   -0.1094
    0.1717    0.0504   -0.0399    0.0227   -0.3945
    0.0813    0.0730    0.0114   -0.2223    0.3806

rint =

    0.2258    0.8988
   -0.5476    0.2565
   -0.3262    0.5032
   -0.5515    0.4557
   -0.7043    0.2429
   -0.2802    0.6216
   -0.8377    0.1550
   -0.6260    0.4844
   -0.4749    0.4543
   -0.6400    0.4211
   -0.3311    0.6745
   -0.4907    0.5915
   -0.5938    0.5140
   -0.4991    0.5445
   -0.8701    0.0812
   -0.4169    0.5795
   -0.0879    0.2338
   -0.4987    0.5214
```

```
-0.6676    0.2231

-0.0071    0.7682

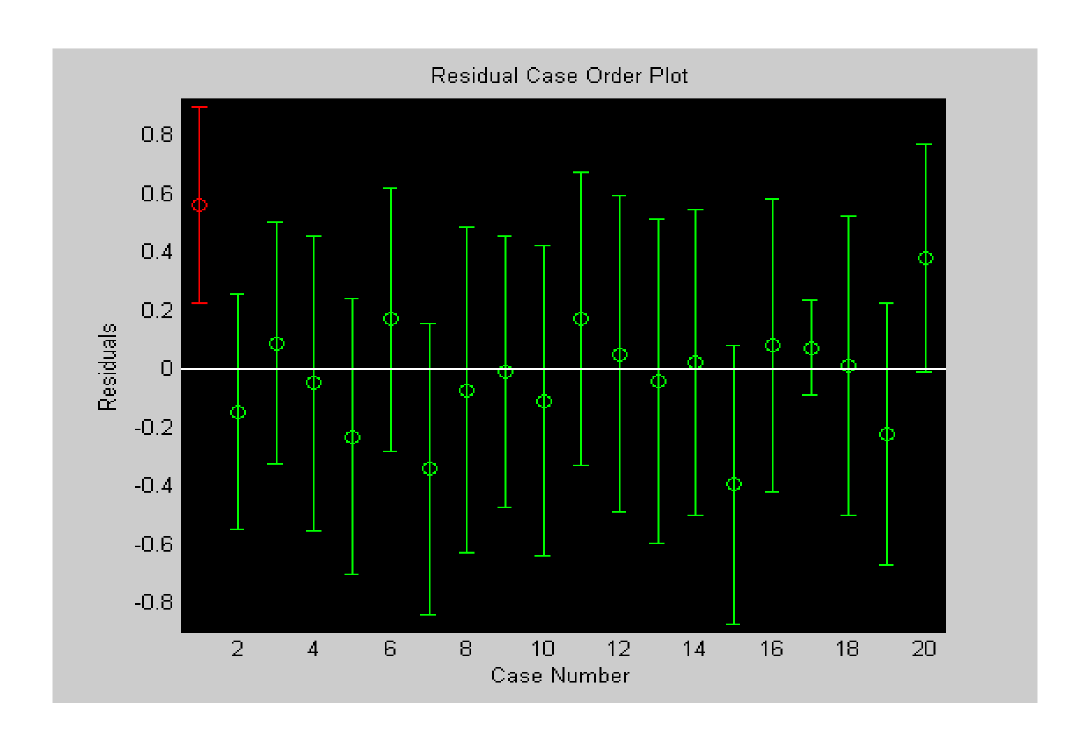

stats =


0.8107   11.9886    0.0001
```

المتغير b يمثل فترات الثقة للمتغير bint وأن المتغير r يمثل المتجه المتبقي rint . توافق المتغير y هو b(1) والذي يقابل فهرسة العمود للواحدات من الاعمدة. عناصر المتجهة stats هو أحصاء الانحدار r2 والاحصاء F وقيمة p وذات العلاقة مع الاحصاء F .

R2 هو معامل التوضيح ويمثل أمكانية المتغيرات المستقلة التي أستطاعت توضيح من المتغير y التابع وفي مثالنل السابق كان معامل التوضيح يساوي ٠.٨١٠٧ وهي تدل على نسبة توضيح عالية بلغت ٨٠%. وأن قيم المعالم المتبقية في النموذج كانت قريبة من الصفر وهذا يدل على أن هذه المتغيرات كان تأثيرها ضعيف جدا أي أن عوامل معاملات الانحدار مساوية ويمكن رسمها كما في الشكل ٨.٣ وذلك بأستخدام الدالة rcoplot وكما يلي:

```
>> rcoplot ( r, rint )
```



شكل ٨.٣ رسم مخرجات الانحدار في المثال السابق rcoplot بأستخدام الدالة

الشكل ٨.٣ يوضح أن ٩٥% من عوامل الثقة رسمت كأعمدة خطأ، حيث أن الاستنتاج الاول يوضح أن عمود الخطأ لا يعبر خط نقطة الصفر الدالة. بأستخدام الدالة التي ستعرض في الفقرة اللاحقة وهذه الدالة بالامكان أن تشكل المصفوفة المطلوبة مع قيم المتنبئات وقيم مربعاتها وقيم مكعباتها وهكذا.

٨.٣.٣ برنامج توافق المنحنى المتكرر **Polynomial Curve Fitting Demo**

البرنامج الجاهز polytool هو بيئة الرسم التفاعلي لبرنامج توافق المنحنى المتكرر والتنبأ حيث بامكانك استخدام الدالة polytool لعمل توافق المنحنى والتنبأ لأي مجموعة بيانات x- y ولكن من اجل عرض الحزمة الاحصائية تجهز مجموعة البيانات polydata.mat

وذلك لعرض بعض المفاهيم الأساسية. ومع هذا البرنامج الجاهز polytool يمكنك عمل ما يلي :

١- رسم البيانات وتوافق التكرار ونقاط الثقة العامة في القيمة المتنبأة الجديدة .

٢- تغيير درجة توافق التكرار .

٣- تقييم التكرار كقيمة x الخاصة .

٤- عرض قيمة y المتنبأة وغير متأكدة في قيمة x الحالية.

٥- السيطرة على نقاط الثقة والاختيار بين المربعات الدنيا أو التوافق .

٦- ترحيل النتائج المتوافقة للتطبيق.

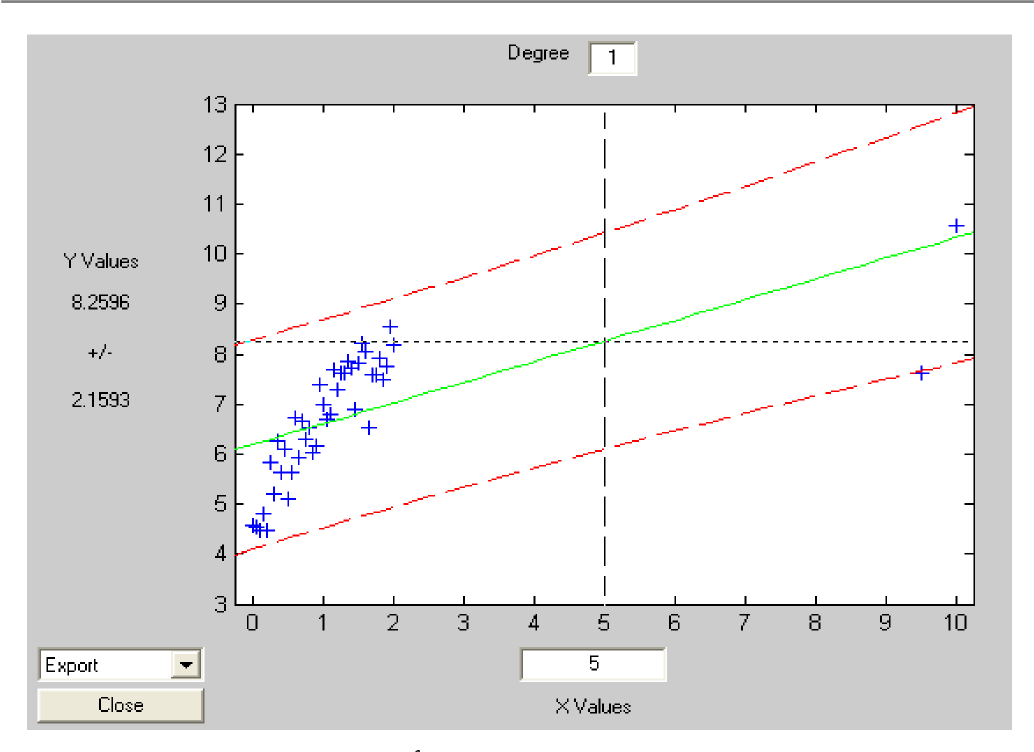
يتضمن توافق التكرار ويمثل كما يلي: Polytoolوأستخدام الدالة

١- تحميل البيانات حيث أنه يجب أولا تحميل البيانات فقبل البدأ بالعرض وهذا المثال يستخدم الملف polydata .mat وهذه المجموعة فان المتغيرات y ، x تستنتج على أنها موجودة مع خطأ في التكرار الثلاثي وان المتغيرات x1 و y1 هي نقاط البيانات الحقيقية من دالة " true " بدون خطأ وكما يلي :

```
>> load polydata
```

٢- محاولة التوافق المستقيم حيث ننفذ الدالة Polytool ونجهزها مع البيانات ليكون التكرار متوافق وبسبب كون البرنامج لا يحقق درجة التكرار حيث أن دالة Polytool تعطي توافق خطي الى البيانلت وهذا موضح في الشكل ٨.٤ حيث يمثل الرسم التنبؤي للنموذج الخطي ويكون الامر كما يلي:

```
>> polytool(x,y)
```



شكل ٨.٤ الرسم التنبؤي للنموذج الخطي بأستخدام الدالة polytool

٨.٤ نماذج سطح الاستجابة التربيعي Quadratic Response Surface Models

طريقة سطح الاستجابة هي أداة لفهم العلاقة الكمية بين متغيرات متعددة الادخال وبين متغير احادي الاخراج. نأخذ الاخراج z كدالة تكرارية لادخالين هما x ، y حيث أن الدالة $z = f\ (x,y)$ تصف سطح ذو بعدين في الحيز (x,y,z) وبشكل عام بامكانك ان تدخل عدة متغيرات للإدخال وكما ترى وان النتيجة هو سطح معين لكل واحد منهم. فمثلا للإدخال الثلاثة (x1,x2,x3) فان معادلة سطح الاستجابة التربيعي تكتب كما يلي :

$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + ..$ (linear term الحد الخطي)

$+ b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + ..$ (Interaction term الحد التفاعلي)

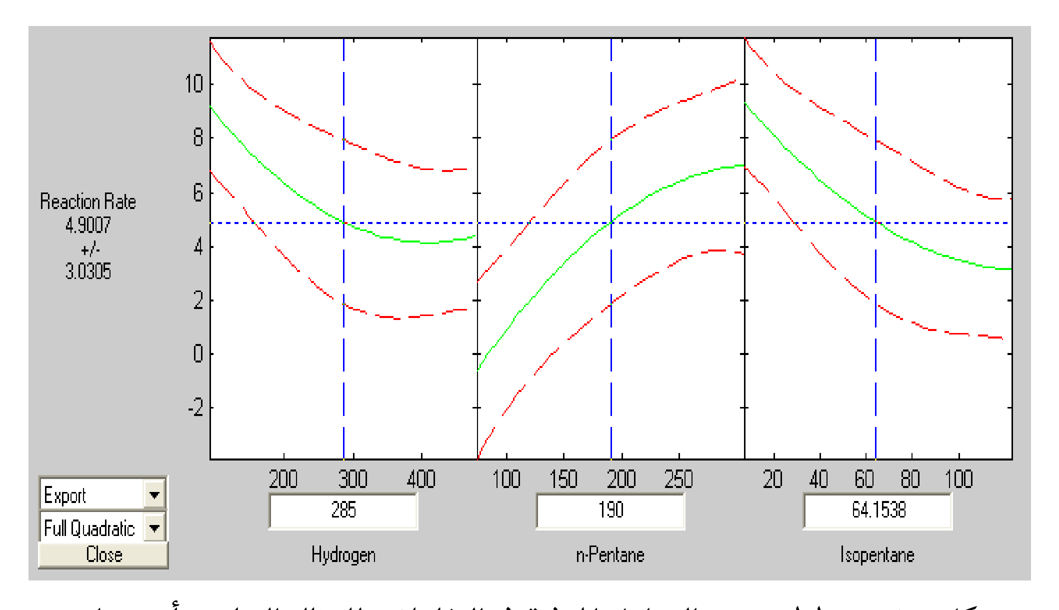
$+ b_{11}x_1^{\ 2} + b_{22}x_2^{\ 2} + b_{33}\ x_3^{\ 2} +..$(quadratic term الحد التربيعي)

الدالة rstool تمثل توافق التفاعل والرسم لسطح الاستجابة متعدد الابعاد وبشكل عام فان واجهة المستخدم عندما تكون على شكل رسومات فإنها تجهز بيئه لعرض مخطط التكرار متعدد الابعاد . ومثال على ذلك نحمل الملف reaction. mat والذي يحتوي على التفاعلات الكامنة كدالة للضغط الجزئي لثلاث مفاعلات كيماوية وهي hydrogen و n-pentane و iso-pentane وكما يلي :

```
>> lood rection
```

```
>> rstool ( reactants , rate , ‘quadratic' , 0.01 , xn , yn )
```

الدالة rstool تعرض المتجه بثلاث رسومات أو منحنيات هي الأول hydrogen وهو متغير مستقل والثاني والثالث هما n-pentane و iso-pentane على التوالي وكما مبين في الشكل 8.5 .



شكل ٨.٥ مخطط يوضح العوامل المؤثرة في التفاعلات للمثال السابق بأستحدام الدالة rstool

٨.٥ الانحدار المرحلي (الخطوات المتعاقبة) Stepwise Regression

الانحدار المرحلي هي تقنيه لاختيار المتغيرات لإدخالها في نموذج متعدد الانحدار. الأنحدار المرحلي الأمامي يبدأ بدون حدود النموذج وفي كل خطوة يضيف حد ذو قيمه إحصائية عليا (بمعنى أخر عندما تكون إحصاء F أعلى وقيمه P أدنى) إلى النهاية. أما الانحدار المرحلي الخلفي فإنه يبدأ بجميع الحدود في النموذج ويحذف الحدود ذات القيمة الأدنى إلى كل المجاميع الفرعية لجميع الحدود وبعدها يضيف الحدود المؤثرة أو يحذف الحدود عديمة التأثير. الحزمة الإحصائية تشمل دالتين لتمثيل الانحدار المرحلي هي :

١. دالة stepwise وهي أداة تفاعل مخططي تجعلك قادر على عرض الانحدار المرحلي .

٢. دالة stepwisefit وهي أداه لتكوين الانحدار المرحلي وباستخدامها بإمكانك إرجاع نتائج الانحدار المرحلي إلى بيئة عمل MATLAB .

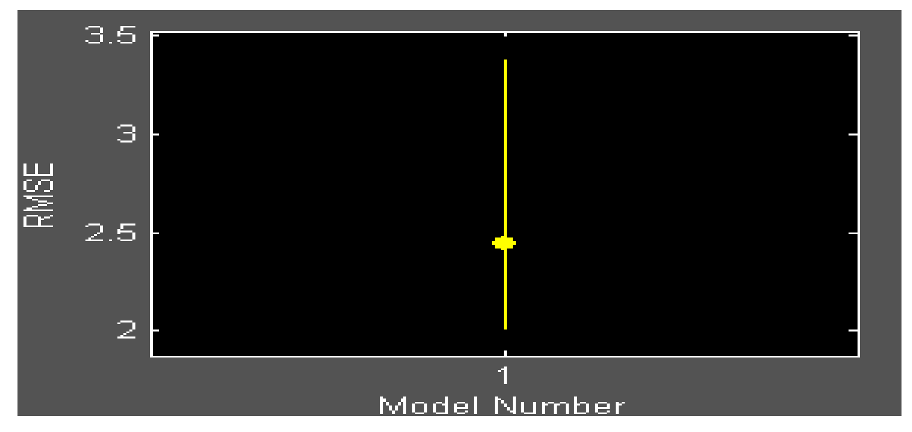
دالة الانحدار المرحلي تعطي واجهة التفاعل التخطيطي والتي تستخدم لمقارنه النماذج التنافسية، وكمثال على ذلك نستخدم ملف البيانات hald والذي يمثل دراسة حرارة التفاعل لمختلف الأمزجة الأسمنتية حيث توجد هنا أربع مركبات في كل مزيج وكمية الحرارة المتولدة تعتمد على الكمية لكل جزء من المزيج وهذا موضح في الاشكال ٨.٦، ٨.٧، ٨.٨، ولتحميل الدالة وتنفذها نتبع ما يلي :

```
>>Load hald
>> stepwise ( ingredients, heat  )
```

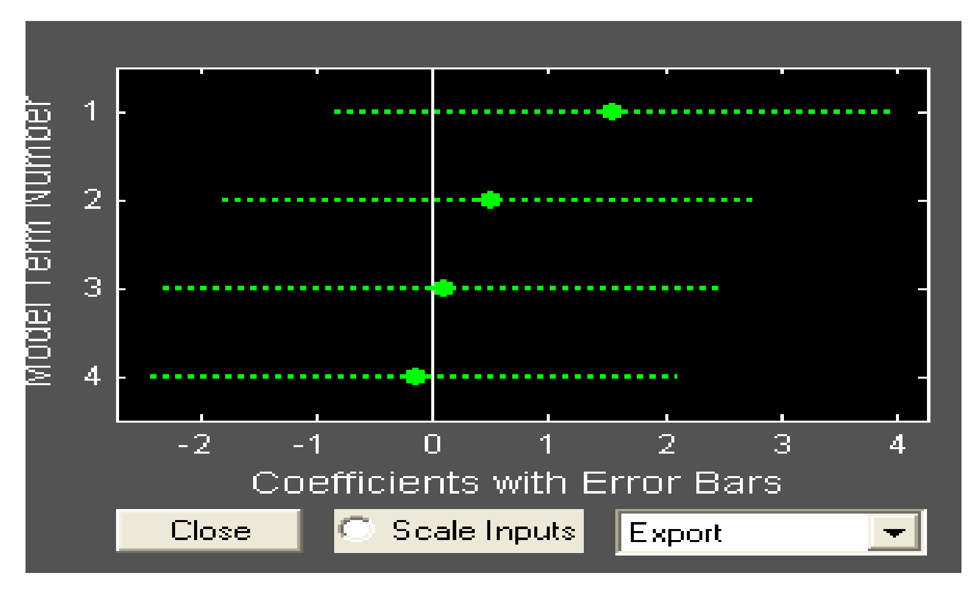
Column #	Parameter	Confidence Intervals Lower	Upper
1	1.551	-0.8319	3.934
2	0.5102	-1.806	2.826
3	0.1019	-2.313	2.517
4	-0.1441	-2.413	2.125
RMSE	R-square	F	P
2.446	0.9824	111.5	4.756e-007

Close Help

شكل ٨.٦ جدول الانحدار المرحلى للملف hald في المثال السابق



شكل ٨.٧ مخطط يبين نسبة الخطأ للملف في المثال السابقhald



شكل ٨.٨ مخطط يبين عوامل نسبة الخطأ للملف hald في المثال السابق

لكل حد في المحور y فان الرسم يوضح الانحدار (المربعات الدينا) للعوامل كنقاط والأعمدة الأفقية تمثل فترات الثقة . النقاط في الشكل تمثل الحدود في النموذج أما النقاط الحمراء فتمثل الحدود التي ليست حاليا ضمن النموذج والأعمدة الأفقية تمثل ٩٠% ملونة و ٩٥% غير ملونة وهي تمثل فترات الثقة. يوجد بالجانب الأيمن لكل عمود حيث أن الجدول يعطي قيمة عامل الانحدار لذلك الحد مع إحصاء t وقيمة p . أن عامل الحد الذي لا يكون في النموذج هو العامل الذي ينتج من إضافة الحد إلى النموذج الحالي. من واجهة الانحدار المرحلي يتم اختبار مقياس للإدخال لتركيز وتحديد الأعمدة لداله الإدخال ليكون الانحراف لمعياري يساوي واحد ، وعندما تستخدم داله الانحدار المرحلي بإمكانك أن تمثل الحالة الابتدائية للنموذج وفترات الثقة التي تستخدم .

الدوال في هذا الحقل لها علاقة مع النماذج التي لها علاقة خطية بين الاستجابة واحد أو أكثر من المتنبأت حيث انه في بعض الأحيان ربما تكون العلاقة غير خطية .

لفهم النماذج الخطية العامة نبدأ بأنه كل هذه النماذج لها ثلاثة خواص هي :

١. الاستجابة لها توزيع طبيعي مع متوسط μ.

٢. متجهة العامل يعرض علاقة خطية x * b للمتغير المستقل x .

٣. النموذج يساوي الاثنين معا μ=x* b .

في النماذج الخطية العامة هذة الخواص تعمم كما يلي :

١. الاستجابة لها توزيع ربما يكون طبيعي , normal , binormal ، poisson , gamma , inverse Goussian مع عناصر تحتوي على متوسط μ .

٢. متجه العامل b يعّرف علاقة خطية x * b للمتغير المستقل x .

٣. دالة الربط f (.) تعرف الربط بين الاثنين f (μ) = x * b .

٨.٧ تطبيق على النماذج الخطية العامة Example of generalized linear model

نستخدم البيانات المخزونة في الملف carbig والتي فيها معلومات عن السيارات وأوزانها المختلفة.يتم تسجيل إعداد السيارات لكل وزن والرقم الذي يمثل أقل مسافة تقطعها السيارة للغالون الواحدمن الوقود والتي هي أقل من الهدف على افتراض أنك لاتعرف عدد الأميال التي تقطعها السيارة لكل غالون وفقط يتم تحديد الرقم الذي يجتاز الاختبار. من المعقول فرض أن القيمة تكون ضعيفة تتبع التوزيع الطبيعي مع العنصر N الذي يمثل العدد الكلي مع المتغير p والذي يعتمد على وزن السيارة. الشكل ٨.٩ يوضح أن العلاقة بالنسبة

إلى السيارات ذات القيمة الضعيفة تتبع العلاقة غير الخطية بشكل S ويتم تطبيق الاوامر كما يلي:

```
>> w = [ 2100 2300 2500 2700 2900 3100 3300 3500 3700 3900 4100
4300];

>>  poor = [ 1  2  3  8  8  14  17  19  15  17  21];

>>  total = [ 98  42  31  34  31  21  23  23  21  16  17  21];

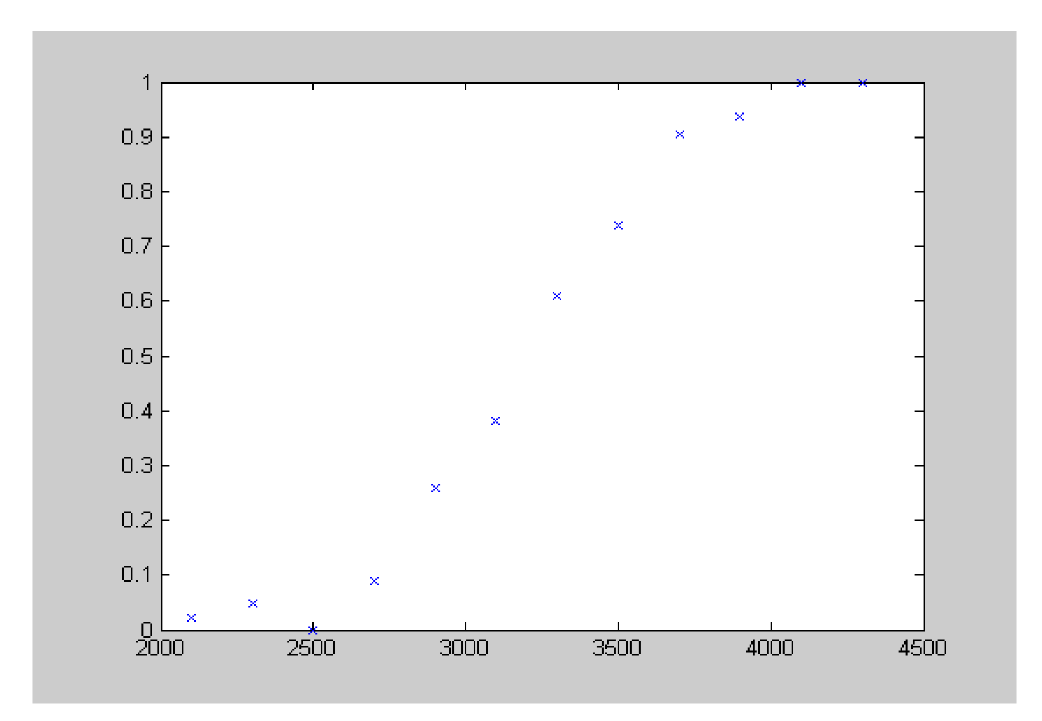
>> [w  poor  total]

ans =

    2100     1     48
    2300     2     42
    2500     0     31
    2700     3     34
    2900     8     31
    3100     8     21
    3300    14     23
    3500    17     23
    3700    19     21
    3900    15     16
    4100    17     17
    4300    21     21
```

ولرسم المتغيرات مع بعضها كما موضحة في الشكل ٩.٩ نكتب أمر الرسم كما يلي:

```
>> ptot (w, poor./total,' x')
```



شكل ٨.٩ رسم بياني للمتغيرات في المثال السابق وربطها مع بعضها

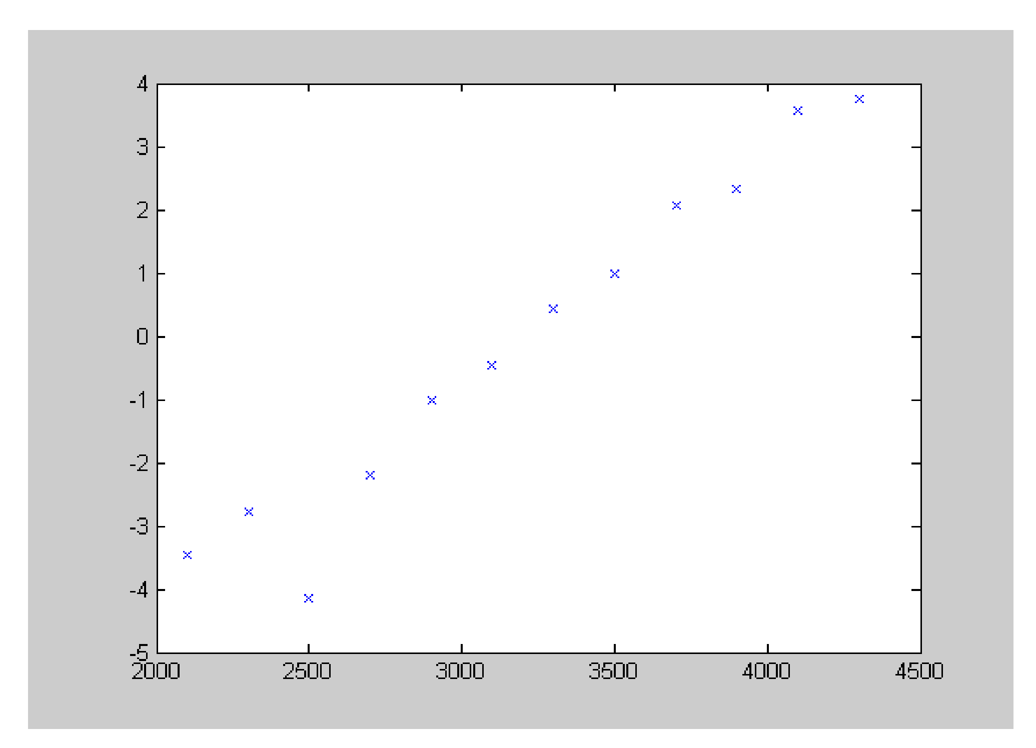
الشكل ٨.٩ يمثل علاقات الرسومات الاعتيادية وله حدود طبيعية في (0.0 و1.0) . نموذج الانحدار الخطي لا ينتج توافق مقنع لهذا المنحي أي أن الخط لا يتوافق مع نقاط البيانات إذ يولد علاقات غير حقيقة اقل من ٠.٠ للسيارات الخفيفة وأكبر من 1.0 للسيارات الثقيلة . توجد أصناف من نماذج الانحدار لتمثيل البيانات ذات العلاقة ومن هذه النماذج هو Logistic Model وهو يعرف العلاقة بين علاقة P والوزن w وكما في المعادلة :

$$Log (p/(1-p)) = p_1 + p_2 w$$

على أية حال بعض العلاقات هي 0 و ١ لذلك لا يمكن تقييم الطرف الأيسر من المعادلة وأن الخطوة المفيدة هي حساب العلاقات المعدلة بإضافة زيادة قليلة إلى القيم الضعيفة والقيم الكلية مثلا نصف الاستنتاج إلى القيم الضعيفة و استنتاج كامل إلى القيم الكلية وهذا يحفظ العلاقات ضمن الحدود وكما موضح في الشكل ٨.١٠ وتكتب الاوامركما يلي :

```
>> padj = ( poor +0.5 ). / ( total +1) ;

>> plot = (w, log ( padj ). / ( 1- padj ) , " x " )
```



شكل ٨.١٠ رسم بياني يوضح التغيير في أوزان القيم في المثال السابق

يمكنك ان تستخدم الدالة glmfit لتوفيق هذا النموذج وكما يلي :

```
>> b = glmfit ( w, [ poor total ], 'binomial' )

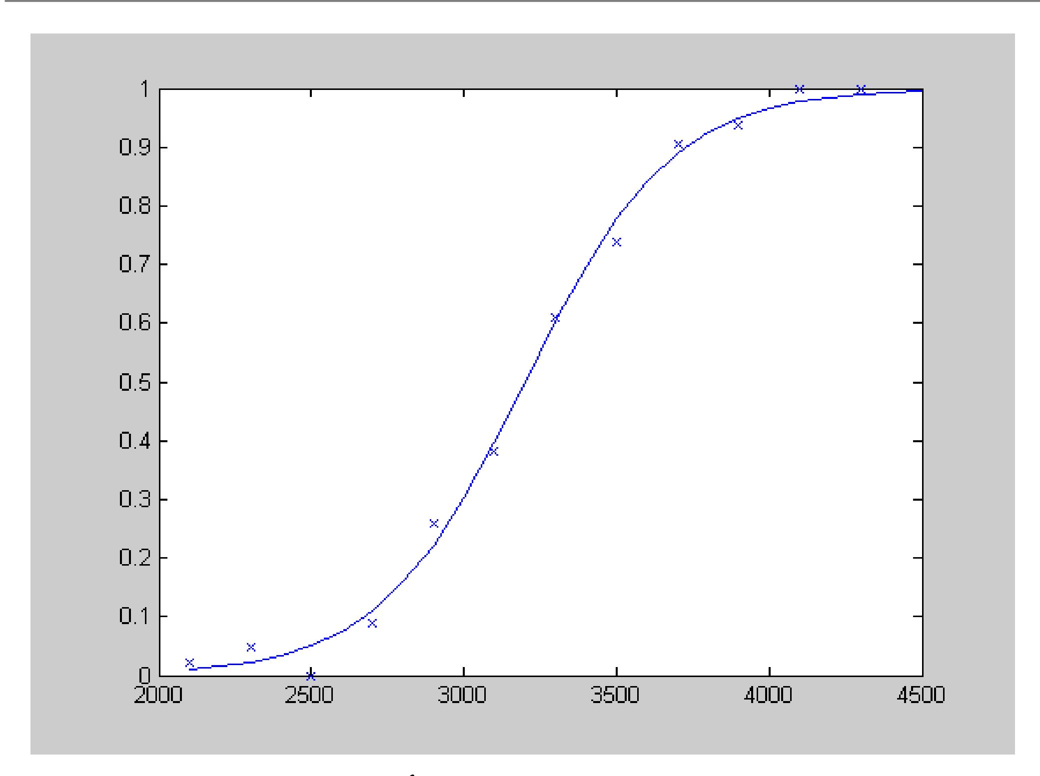
b = - 13.3801            .0042
```

لاستخدام هذه العوامل لحساب نسبة التوافق ، يجب أن تقلب العلاقة وفقا للعلاقات الرياضية لتكوين المعادلة حيث تكون بشكل مفيد وأن الداله glmval بإمكانها أن تكتب داله الربط هذه لحساب القيم المتوافقة وباستخدام هذه الداله بالإمكان رسم علاقات التوافق لمدى أوزان السيارات وباستغلال أعلى لهذا المنحى ذو الرسم المشتت الأصلي وكما موضح في الشكل ٨.١١ وتكتب الاوامركما يلي :

```
>> x = 2100 : 100 : 4500;

>> y = glmval ( b, x, 'logit');

>> ptot ( w, poor . / total , 'x',x , y ,'r-' )
```



شكل ٨.١١ الرسم المشتت للمثال السابق بأستخدام الدالة glmval

النماذج الخطية العامة يمكن أن توافق توزيعات مختلفة وبعلاقات مختلفة بين عناصر التوزيع والمتنبئات. البرنامج الجاهز glmdemo يبدأ بعرض شرائح تصف النماذج الخطية العامة والتي تعرض أمثله لأي من الدوال أو التوزيعات والموجودة مع النماذج الخطية العامة.

٨.٨ الطرق اللامعلمية النشيطة Robust and Nonparametric Methods

الحقل الإحصائي في حزمه MATLAB لها داله ارتداد قوية وهي مفيدة عندما تكون خارج، وان الطرق القوية صممت لتكون مكثفة نسبياً للمتغيرات الكبيرة في الجزء القليل من البيانات. الحقل الإحصائي كذلك لها نسخ nonparametric لتحليل دوال التباين ذات الطريق الواحد وذات الطرقتين. الاختبارات التقليدية غير مرغوبة والاختبارات الـ nonparametric تعمل فقط على افتراض معتدل للبيانات(كما في التوزيع الطبيعي لحد

الخطأفي بعض الأحيان هذاالافتراض ليس مضمون إذا كان توزيع الخطأ غيرمتناسق أي أن هناك تجاوز لافتراض الأخطاء الطبيعية) ومناسبة عندما يكون التوزيع طبيعي.من جهة أخرى هي أقل قوة من الطرق التقليدية للبيانات الموزعة طبيعياً مثال على ذلك نأخذ الانحدار الخطي المتعدد والذي يدل على أنه يوجدخارج عندما تستخدم ارتدادالمربعات الصغرى الاعتياديةلنمذجةالاستجابةكداله لخمسة متنبئات . ويوضح المثال التالي كيفية تخمين العوامل باستخدام الدالةrobustfit حيث نحمل الملف moor ونطبق الاوامركما يلي:

```
>> load  moore
>> = moore( : ,1 :5) ;
>>moore (: , 6) ‘
>> ( br, statsr) = robustfit (x , y) ;
>> br
br =  -1.7742
         0.0000
         0.00009
         0.0002
         0.0062
         0.0001
```

نقارن هذه التخمينات مع التي حصلت من دالة التباين

```
>> b
b = - 2.1561
        - 0.0000
 0.0013
 0.0001
 0.0079
 0.0001
```

لفهم الاختلاف بين الاثنين فمن المفيد ملاحظة متغير الوزن من توافق robust وهو يقيس كيف ، أن كمية الوزن أعطيت لكل نقطة خلال التوافق وفي هذه الحالة القيمة الأولى لها وزن قليل جداً لذلك يهمل تأثيرها وكما يلي :

```
>> statsr . w'

ans =

0.0577    0.9977    0.9776    0.9455    0.9687

0.8734    0.9177    0.9990    0.9653    0.9676

0.9768    0.9882    0.9998    0.9979    0.8185

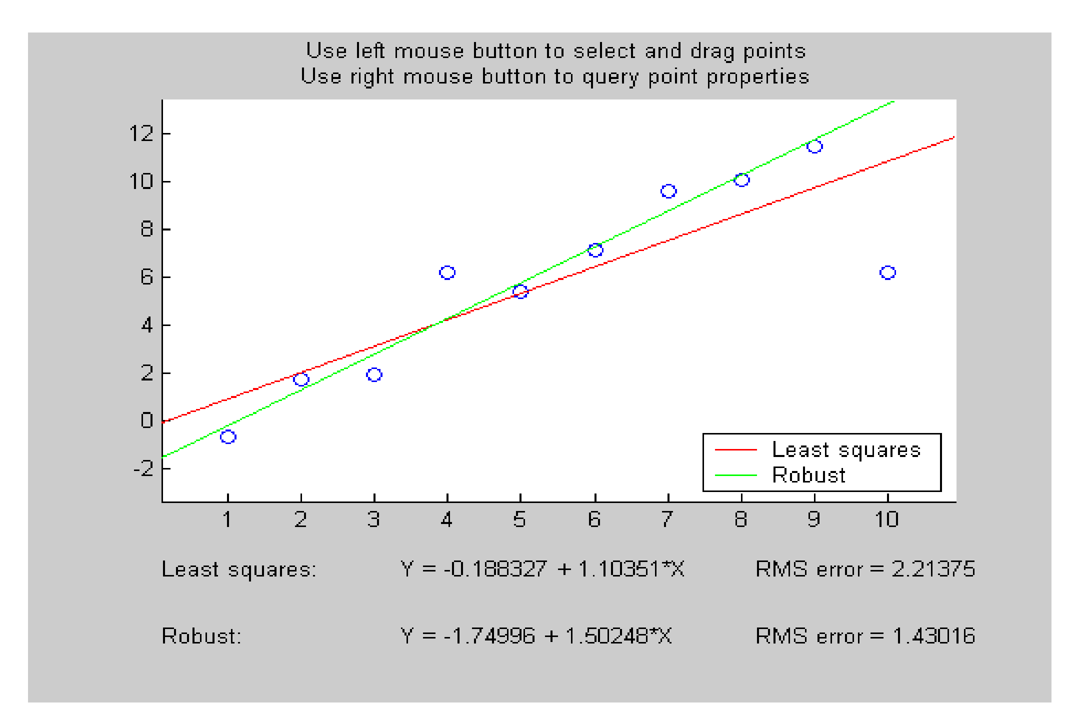
0.9757    0.9875    0.9991    0.9021    0.6953
```

البرنامج الجاهز robustdemo يمثل مقارنه بسيطة للمربعات الصغرى لوتوافق robust الاستجابة ومتنبأ واحد حيث بالامكان استخدام بيانات تجهز بالبرنامج وكذلك بأمكانك تجهيز البيانات الخاصة وهذه الدالة تمثل بالخطوات التالية :

١. بداية البرنامج باستخدام البرنامج الجاهز robustdemo بناء نموذج البيانات وذلك بإدخال أسم الدالة وكما يلي:

```
>> robustdemo
```

النتيجة تمثل رسم مشتت مع خطين متوافقين احدهما متوافق مع ارتداد المربعات الصغرى الاعتيادية والأخر متوافق مع ارتداد robust . في نهاية الشكل توجد المعادلات للخط المتوافق وخطأ الانحراف المعياري المخمن لكل توافق . تأثير أي نقطة لتوافق المربعات الصغرى يعتمد على المتبقى والفائدة لتلك النقطة ، حيث أن المتبقى هو المسافة العمودية من النقطة الى الخط وأما الفائدة فهي قياس كم تبعد النقطة من مركز البيانات x. كذلك فأن تأثير أي نقطة على توافق robust ، يعتمد على الوزن الذي يرمز إلى النقطة وان النقاط البعيدة من الخط تأخذ أقل وزن وكما موضحة في الشكل ٨.١٢.

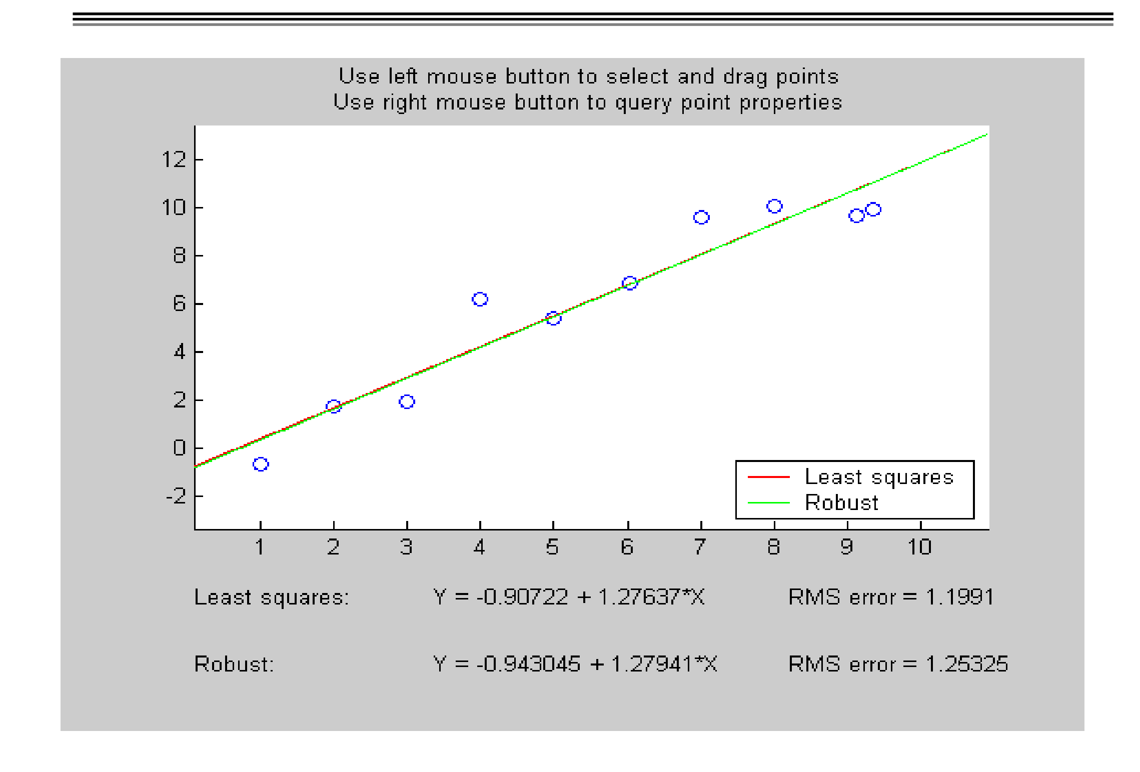


شكل ٨.١٢ المنحنيات للطريقة النشيطة والمربعات الصغرى في البرنامج robustdemo

٢. مقارنه تأثير الفائدة والوزن وذلك باستخدام الطرف الأيمن للفأر لتأشير على أي نقطة وملاحظة المربعات الدنيا للفائدة وأوزان الـ robust .

في هذا المثال أن نقطة rightmost لها قيمه فائدة ٠.٣٥ وهي كذلك تبعد عن الخط لذلك فهي تبذل تأثير كبير لتوافق المربعات الدنيا ولها وزن صغير لذلك تستبعد تماما من توافق robust.

٣. ملاحظة كيف أن التغيرات في البيانات يؤثر التوافقين وذلك باستخدام الطرف الأيسر للفأر حيث يتم اختبار أي نقطة وتسحب إلى الموقع الجديد بينما يبقى الطرف الأيسر للفأر إلى الأسفل وكما موضح في الشكل ٨.١٣ .



شكل ٨.١٣ المنحنيات في الشكل ٨.١٢ بعد أجراء بعض التعديلات عليها

المصادر

References on Statistic المصادر في مجال الأحصاء

1. Aliaga & Gunderson, "Interactive Statistics", 3/E, 2006, Prentice Hall.

2. Agresti & Franklin, "Statistics: The Art and Science of Learning From Data", 2007, Prentice Hall.

3. Bickel & Doksum, "Mathematical Statistics, Updated Printing", 2/E, 2007, Prentice Hall.

4. Cody & Smith, "Applied Statistics and the SAS Programming Language", 5/E, 2006, Prentice Hall.

5. Dretzke, "Statistics with Microsoft Excel", 3/E, 2005, Prentice Hall.

6. Freund, "Modern Elementary Statistics", 11/E, 2004, Prentice Hall.

7. Freund & Perles, "Statistics: A First Course", 8/E, 2004, Prentice Hall.

8. Freund & Perles, "Modern Elementary Statistics", 12/E, 2007, Prentice Hall.

9. Hair, Black, Babin, Anderson & Tatham, "Multivariate Data Analysis", 6/E, 2006, Prentice Hall.

10. Hogg, Craig & McKean, "Introduction to Mathematical Statistics", 6/E, 2005, Prentice Hall.

11. Johnson & Wichern, "Applied Multivariate Statistical Analysis", 5/E, 2002, Prentice Hall.

12. Larsen & Marx, "Introduction to Mathematical Statistics and Its Applications", An, 4/E, 2006, Prentice Hall.

13. Larson & Farber, 'Elementary Statistics: Picturing the World", 3/E, 2006, Prentice Hall.

14. Levine & Stephan, "Even You Can Learn Statistics: A Guide for Everyone Who Has Ever Been Afraid of Statistics", 2005, Prentice Hall.

15. McClave, Benson & Sincich, "Statistics for Business and Economics", 9/E, 2005, Prentice Hall.

16. McClave & Sincich, "First Course in Statistics", A, 9/E, 2006, Prentice Hall.

17. McClave & Sincich, "Statistics", 10/E, 2006, Prentice Hall.

18. Mendenhall & Sincich, "Statistics for Engineers and the Sciences", 5/E, 2007, Prentice Hall.

19. Mendenhall & Sincich, "Second Course in Statistics, A: Regression Analysis", 6/E, 2003, Prentice Hall.

20. Pursley, "Random Processes in Linear Systems" 2002, Prentice Hall.

21. Sincich, Levine & Stephan, "Practical Statistics by Example Using Microsoft Excel and Minitab", 2/E, 2002, Prentice Hall.

22. Samuels & Witmer, "Statistics for the Life Sciences", 3/E, 2003, Prentice Hall.

23. Sullivan, "Fundamentals of Statistics", 2005, Prentice Hall.

24. Sullivan, "Statistics: Informed Decisions Using Data", 2/E, 2007, Prentice Hall.

25. Tamhane & Dunlop, "Statistics and Data Analysis: From Elementary to Intermediate", 2000, Prentice Hall.

26. Walpole, Myers, Myers & Ye, "Probability & Statistics for Engineers & Scientists", 8/E, 2007, Prentice Hall.

References on MATLAB   MATLAB المصادر في مجال حزمة

1. Breiner & Biran, "MATLAB 6 for Engineers", 2002, Prentice Hall.

2. Etter, Kuncicky & Moore, "Introduction to Matlab 7", 2005, Prentice Hall.

3. Hanselman & Littlefield, "Mastering MATLAB 7: International Edition", 2005, Prentice Hall.

4. Kuncicky, "MatLAB Programming", 2003, Prentice Hall.

5. MathWorks, Inc., "MATLAB and Simulink Student Version Release 14", 2006, Prentice Hall.

6. Moore, "MATLAB for Engineers", 2007, Prentice Hall.

7. Magrab, Azarm, Balachandran, Duncan, Herold & Walsh, "Engineer's Guide to MATLAB", 2/E, 2005, Prentice Hall.

8. Recktenwald, "Introduction to Numerical Methods and MATLAB: Implementations and Applications", 2001, Prentice Hall.